



Séminaires pour le calibrage des productions orales
écrites illustrant les niveaux A1.1 et A1 du *Cadre
européen commun de référence pour les langues*

CIEP, Sèvres, 29 mai et 11 juillet 2006

Rapport d'analyse

Cellule qualité et expertises

Pôle évaluations et certifications

Centre international d'études pédagogiques

Rapport d'analyse :

Calibrage des productions orales - Séminaire du 29 mai 2006 : page 3 / 110

Calibrage des productions orales - Séminaire du 11 juillet 2006 : page 50 / 110

Calibrage des productions orales - Séminaire du 29 mai 2006 : page 85 /110



Calibrage des productions orales

Séminaire du 29 mai 2006

Rapport d'analyse

Cellule qualité et expertises

Pôle évaluations et certifications



Sommaire

INTRODUCTION.....	3
1. LES DONNEES	3
1.1. LES PARTICIPANTS	3
1.2. LES PRODUCTIONS ORALES	3
2. ETUDE DE L’EVALUATION DES EXPERTS	3
2.1. DEGRE D’ACCORD ENTRE LES EXPERTS	3
2.2. LA SEVERITE DES GROUPES D’EXPERTS.....	3
2.3. ACCORD PAR NIVEAU DE COMPETENCE.....	3
3. NIVEAUX DES CANDIDATS.....	3
3.1. AHMED.....	3
3.2. AMINA E.	3
3.3. AMINA I.	3
3.4. BELAL	3
3.5. FELIX.....	3
3.6. GANISTA	3
3.7. HALA	3
3.8. MANSOUR	3
3.9. MUNAZZA.....	3
3.10. SALWA.....	3
3.11. YAZID.....	3
CONCLUSION	3
ANNEXE	3
KAPPA POUR PLUSIEURS JUGES ET PLUSIEURS CATEGORIES (FLEISS).....	3
KAPPA – DECRIRE – ETENDUE.....	3
KAPPA –PRESENTER – AISANCE	3
KAPPA –PRESENTER – INTERACTION	3



Introduction

L'objectif du séminaire organisé par le Centre International d'Etudes Pédagogiques et la Délégation générale à la langue française et aux langues de France, était de permettre aux experts réunis à cette occasion d'évaluer, selon les procédures définies par la Division des politiques linguistiques du Conseil de l'Europe, un certain nombre de productions orales de candidats au DILF (Diplôme initial de langue française), afin d'illustrer les différences existant entre le niveau A1.1 tel qu'il est défini dans le *Référentiel pour les premiers acquis en français* et le niveau A1 du *Cadre européen commun de référence pour les langues* (CECR).

Ces productions orales, après avoir été calibrées par la cellule qualité et expertise du CIEP, ont donné lieu à la création d'un DVD (*Productions orales et écrites illustrant, pour le français, les niveaux A1.1 et A1 du Cadre européen commun de référence pour les langues*).



1. Les données

1.1. Les participants

Les participants sont des experts en FLE (Français Langue Etrangère)¹, ils sont 13 au total et pour les besoins du vote, sont numérotés de 1 à 15 (les numéros 5 et 6 n'ayant pas été attribués).

Ils estimeront tout d'abord individuellement le niveau des productions orales des 11 candidats présentés. Dans un second temps, une analyse, par groupes de 3 à 4 personnes, permettra d'obtenir un consensus sur un niveau de compétence précis.

Nous formons 4 groupes composés des linguistes suivants :

- Groupe 1 : linguistes 1, 14 et 15,
- Groupe 2 : linguistes 2, 9 et 10,
- Groupe 3 : linguistes 11, 12 et 13,
- Groupe 4 : linguistes 3, 4, 7 et 8.

1.2. Les productions orales

Onze productions orales ont été présentées aux experts. Deux experts n'en ont notées que 9.

Les productions comportent 3 activités :

- décrire,
- exprimer,
- présenter.

Les experts donnent une note globale à chaque activité et une note pour chacun des critères suivants :

- étendue,
- correction,
- aisance,
- interaction,
- cohérence.

¹ Liste des experts : Gilles BRETON, Jean-Claude BEACCO, Sébastien GEORGES, Patrick Riba, Gilbert LHOTE, Dorothee DUPLEIX, Maud LAUNAY, Mariela DE FERRARI, Sophie ETIENNE, Perrine MICHODET, Johanna PANTHIER, Marie-Françoise ROTTIER, Christine TAGLIANTE

Les experts doivent évaluer le niveau des candidats entre les niveaux A1.1 - et A2, soit :

A1.1-

A1.1

A1.1+

A1

A1+

A2.

Nous étudierons en premier lieu l'homogénéité de la correction des experts, puis le niveau des candidats. Nous ne calibrerons pas les productions orales grâce au modèle de Rasch à multiple facettes car nous ne disposons pas d'un nombre assez important d'experts.



2. Etude de l'évaluation des experts

2.1. Degré d'accord entre les experts

Nous souhaitons savoir si les experts jugent de la même façon le niveau des candidats. Nous utilisons pour mesurer le degré d'accord des experts le coefficient kappa (cf. annexes p. 3).

Les 13 experts ont classé 11 candidats selon leur niveau en français (le niveau en français ici est composé en 6 catégories). Le coefficient kappa k mesure le degré d'accord entre les experts. Le coefficient k est compris entre 0 et 1. En cas d'accord parfait entre les experts $k=1$. Si les experts jugent le niveau des candidats d'une façon aléatoire, non cohérente, $k=0$. Plus le coefficient k est proche de 1, plus le degré d'accord entre les experts sera important.

Nous étudierons l'accord des experts uniquement sur 9 candidats car les juges 3 et 4 n'ont pas noté les candidats Salwa et Yazid. Nous vérifierons si les experts jugent de façon aléatoire (H_0 : " $k=0$ "). Si le test est significatif ($p < 5\%$), nous rejetons cette hypothèse avec un risque d'erreur de 5% (risque de 1^{ère} espèce $\alpha=5\%$).

Valeurs du coefficient kappa

Activités	Critères	Votes individuels			Votes après discussion		
		kappa	z	p	kappa	z	p
DECRIRE	note	0,442	16,20	0	0,735	27,90	0
	étendue	0,229	8,17	0	0,593	22,30	0
	correction	0,388	14,80	0	0,613	23,20	0
	aisance	0,320	13,70	0	0,703	27,40	0
	interaction	0,330	13,50	0	0,640	25,00	0
	cohérence	0,345	13,50	0	0,659	25,30	0
EXPRIMER	note	0,406	15,40	0	0,720	26,90	0
	étendue	0,248	9,40	0	0,614	22,90	0
	correction	0,230	9,04	0	0,600	22,30	0
	aisance	0,324	13,30	0	0,651	24,70	0
	interaction	0,298	11,50	0	0,711	27,00	0
	cohérence	0,358	15,00	0	0,708	26,50	0
PRESENTER	note	0,388	14,30	0	0,768	25,80	0
	étendue	0,294	11,10	0	0,836	27,00	0
	correction	0,314	11,90	0	0,681	22,90	0
	aisance	0,340	14,20	0	0,545	19,70	0
	interaction	0,267	11,10	0	0,545	19,70	0
	cohérence	0,246	10,20	0	0,652	24,40	0

Lors des votes individuels, le degré d'accord des experts est plutôt moyen, voire faible selon les critères. Quels que soient les notes ou les critères observés, le degré d'accord des experts est bien plus élevé après le travail de groupe.

Après le travail de groupe, le jugement des experts est très homogène pour les 3 notes globales et le critère *étendue* de l'activité *présenter* (les coefficients kappa sont très proches de 1). Pour tous les critères, le jugement des experts est assez homogène (les coefficients kappa sont assez proches de 1). Le degré de concordance du critère *étendue* de l'activité *décrire* et des critères *aisance* et *interaction* de l'activité *présenter*, est plutôt moyen.

Le test d'hypothèse des jugements aléatoires est rejeté ($p < 5\%$) pour chaque note globale et chaque critère observé. Les résultats du test indiquent que les experts n'évaluent pas de façon aléatoire les productions orales des candidats.

Pour les trois critères dont le degré de d'accord est plutôt moyen, une analyse de la concordance a été menée en comparant l'analyse des experts deux à deux (cf. annexes p. 3, 3 et 3). L'objectif est de comprendre pourquoi le degré d'accord des experts est plus faible pour ces 3 critères.

Cette étude montre que les experts 1, 14 et 15 ont une façon de noter assez homogène quand on les compare deux à deux. Les 10 autres experts notent également de façon assez homogène. En revanche, les experts 1, 14 et 15 ont un degré d'accord très faible avec les 10 autres. Le test d'hypothèse que deux experts évaluent les candidats de façon arbitraire confirme la non-concordance observée entre les experts du groupe 1 (1, 14 et 15) et les 10 autres experts. Ainsi, les linguistes des groupes 2, 3 et 4 évaluent de façon homogène les candidats. Il est à noter que le jugement de ces 10 linguistes ne concorde pas avec celui des 3 linguistes du groupe 1.

Le jugement des experts est plus homogène après le travail de groupe. Nous nous baserons sur les votes après discussion pour établir le niveau des candidats.

2.2. La sévérité des groupes d'experts

Pendant la discussion, les experts ont travaillé par groupe de 3 à 4 personnes.

L'objectif de l'étude suivante est de repérer les groupes experts qui sont plus sévères ou plus indulgents.

Les candidats sont notés sur 6 niveaux : de A1.1- à A2. Leurs notes sont exprimées par une variable ordinale, nous n'utiliserons donc pas l'ANOVA "classique", mais le test H de Kruskal-Wallis². Nous chercherons à savoir si le fait que les candidats soient évalués par des groupes d'experts différents a impact sur leurs notes. Si le test est significatif ($p < 0.05$), les résultats des candidats sont influencés significativement par le groupe de linguistes qui les ont corrigés. Dans ce cas uniquement, nous calculerons les comparaisons post-hoc des rangs moyens de tous les couples de groupes d'experts, afin de savoir plus précisément quel sont les groupes qui sont plus ou moins sévères.

2.2.1. Décrire

Après discussion, pour la note globale et les 5 critères, nous n'avons pas montré que les notes des candidats étaient significativement différentes quand les candidats étaient évalués par des groupes experts différents (le test H de Kruskal-Wallis n'est pas significatif).

DECRIRE - ANOVA de Kruskal-Wallis

	Votes après discussion
note	H (3, N= 139) =2,045421 p =,563
étendue	H (3, N= 139) =1,887730 p =,596
correction	H (3, N= 139) =4,016423 p =,259
aisance	H (3, N= 139) =,4350156 p =,932
interaction	H (3, N= 139) =2,667922 p =,445
cohérence	H (3, N= 139) =1,966023 p =,579

2.2.2. Exprimer

Après le travail de groupe, pour la note globale et les 4 critères *étendue*, *aisance*, *interaction* et *cohérence*, nous n'avons pas montré que les notes des candidats sont significativement différentes quand elles sont attribuées par des groupes d'experts différents (le test H de Kruskal-Wallis n'est pas significatif).

² Ce test est utilisé pour comparer au moins trois échantillons, et tester l'hypothèse nulle suivant laquelle les différents échantillons à comparer sont issus de la même distribution ou de distributions de même médiane.

Pour le critère *correction*, nous observons des différences significatives entre les votes des 4 groupes de linguistes. Nous appliquons le test de comparaison post hoc, afin de savoir quel groupe est plus sévère que les autres. Le test est significatif pour les groupes 2 et 3. Les linguistes du groupe 2 notent plus sévèrement que ceux du groupe 3.

EXPRIMER - ANOVA de Kruskal-Wallis

	Votes après discussion
note	H (3, N= 139) =3,799779 p =,283
étendue	H (3, N= 139) =7,485118 p =,057
correction	H (3, N= 139) =9,552350 p =,022
aisance	H (3, N= 139) =2,166701 p =,538
interaction	H (3, N= 139) =5,037109 p =,169
cohérence	H (3, N= 139) =5,873076 p =,118

EXPRIMER –Tests post hoc - Correction – Votes après discussion

Groupes	1 (R*:71,81)	2 (R :53,59)	3 (R :81,04)	4 (R :72,92)
1		0,40	1,00	1,00
2	0,40		0,03	0,25
3	1,00	0,03		1,00
4	1,00	0,25	1,00	

* R : Rangs moyens

2.2.3. Présenter

Après le travail de groupe, pour la note globale et les 5 critères, nous n’avons pas montré que les notes des candidats sont significativement différentes quand elles sont attribuées par des groupes d’experts différents (le test H de Kruskal-Wallis n’est pas significatif).

PRESENTER - ANOVA de Kruskal-Wallis

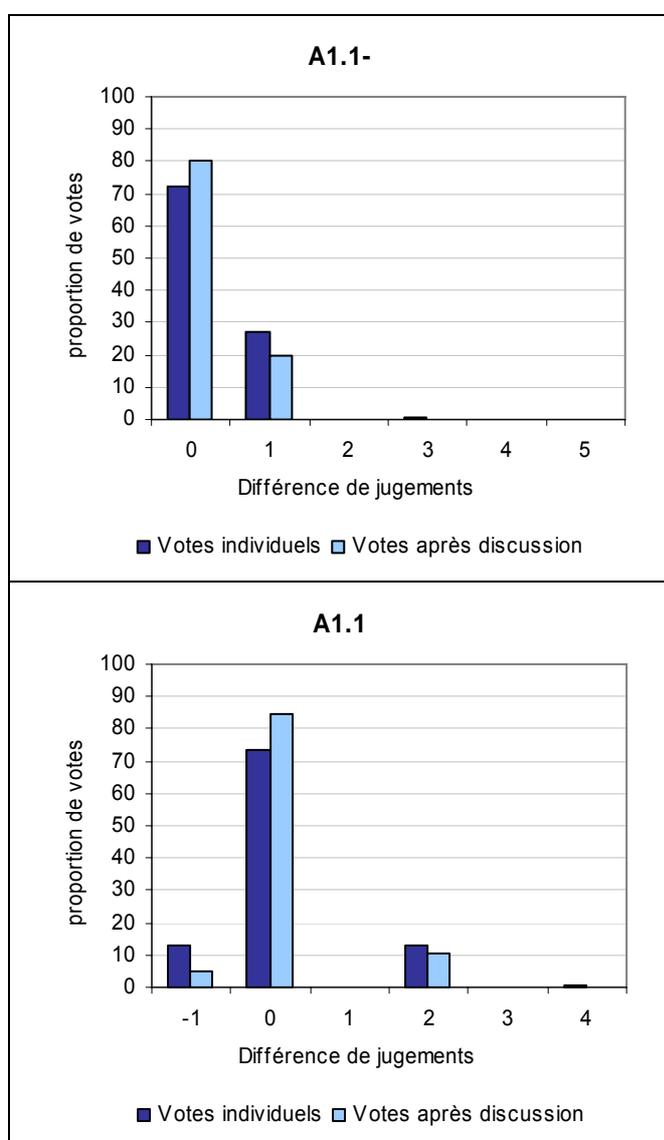
	Votes après discussion
note	H (3, N= 139) =2,903929 p =,406
étendue	H (3, N= 139) =1,354516 p =,716
correction	H (3, N= 139) =4,725900 p =,193
aisance	H (3, N= 139) =1,355188 p =,716
interaction	H (3, N= 139) =,6301290 p =,889
cohérence	H (3, N= 139) =4,786903 p =,188

Pour le critère *correction* de l’activité EXPRIMER, les linguistes du groupe 2 sont plus sévères que ceux du groupe 3. Pour les autres critères et les notes globales, nous n’avons pas montré qu’il existait des différences significatives entre les votes des différents groupes.

2.3. Accord par niveau de compétence

Les graphiques suivants donnent une représentation de l'accord par niveau, avant et après le travail de groupe. Cette analyse est effectuée sur l'ensemble des notes globales et des critères. Cela nous permet, par exemple, de savoir si les experts sont plus souvent en désaccord quand il s'agit de positionner un candidat au niveau A1.1- ou A1.

Pour chaque candidat, nous supposons que son niveau correspond au niveau le plus souvent sélectionné par les experts. Nous observons uniquement 3 niveaux possibles pour les candidats : A1.1-, A1.1 et A1.



0 : les experts sont d'accord, les candidats sont de niveau A1.1-.

1 : les experts pensent que les candidats sont de niveau A1.1.

2 : les experts pensent que les candidats sont de niveau A1.1+.

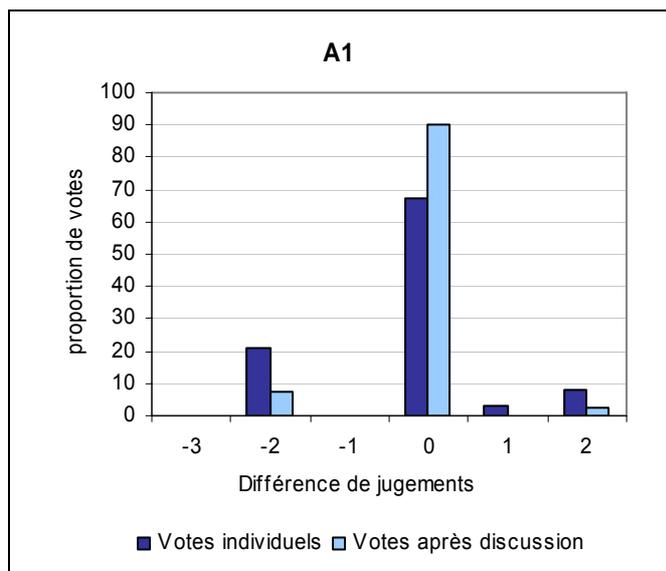
3 : les experts pensent que les candidats sont de niveau A1.1.

-1 : les experts pensent que les candidats sont de niveau A1.1-.

0 : les experts sont d'accord, les candidats sont de niveau A1.1.

2 : les experts pensent que les candidats sont de niveau A1.

4 : les experts pensent que les candidats sont de niveau A2.



-2 : les experts pensent que les candidats sont de niveau A1.1.

0 : les experts sont d'accord, les candidats sont de niveau A1.

1 : les experts pensent que les candidats sont de niveau A1+.

2 : les experts pensent que les candidats sont de niveau A2.

Quel que soit le niveau de compétence des candidats, les experts sont plus souvent d'accord après le travail de groupe.

Lors du vote individuel, ils sont plus souvent d'accord quand les candidats sont estimés de niveaux A1 (74%) ou A1.1 (72%). Le pourcentage d'experts étant d'accord pour évaluer un candidat de niveau A1 est plus faible (67%).

Après discussion, ils sont plus souvent d'accord quand les candidats sont estimés de niveau A1 (90%) que pour les deux autres niveaux (A1.1- : 80% et A1.1 : 84%).

Les experts sont plus souvent d'accord sur le niveau des candidats après le travail de groupe.

3. Niveaux des candidats

Nous rappelons que Yazid et Salwa ont été évalués par 11 experts alors que les 9 autres candidats ont été évalués par 13 experts.

Nous nous baserons sur les votes des experts après le travail de groupe pour évaluer le niveau des candidats. Pour chaque candidat, nous estimerons que son niveau correspond au niveau le plus souvent sélectionné par les experts.

3.1. AHMED

3.1.1. Décrire

Lors du vote individuel, plus de la moitié des experts, 9 sur 13, pensent que ce candidat est de niveau A1. Le niveau d'Ahmed varie suivant les critères observés. L'avis des linguistes est partagé entre les niveaux A1 et A1.1 pour le critère *étendue*. Ils estiment qu'il a le niveau A1 aux critères : *aisance*, *interaction* et *cohérence* et le niveau A1.1 au critère *correction*.

Après discussion en petits groupes, la tendance s'inverse, plus de la moitié des experts, 9 sur 13, pensent qu'Ahmed serait de niveau A1.1. Ils estiment qu'il a le niveau A1.1 aux critères *étendue* et *correction* ; et le niveau A1 aux critères *aisance* et *interaction*. Les votes sont partagés entre les niveaux A1.1 et A1 pour le critère *cohérence*.

3.1.2. Exprimer

Avant et après la discussion, quel que soit le critère observé, la majorité des experts estiment que le candidat est de niveau A1.

3.1.3. Présenter

Que ce soit avant ou après le travail de groupe, la majorité des experts estiment que le candidat est de niveau A1 globalement et pour chaque critère.

Niveaux du candidat par activité

Activité	Niveau
Décrire	A1.1
Exprimer	A1
Présenter	A1

Résultats des votes des experts

AHMED	Vote	Critère	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
DECRIRE	individuels	note	0	4	0	9	0	0	13
		étendue	0	6	0	7	0	0	13
		correction	0	9	0	4	0	0	13
		aisance	0	3	0	8	0	2	13
		interaction	0	2	0	9	0	2	13
		cohérence	1	4	0	7	0	1	13
	après discussion	note	0	9	0	4	0	0	13
		étendue	0	8	0	5	0	0	13
		correction	0	9	0	4	0	0	13
		aisance	0	3	0	9	0	1	13
		interaction	0	4	0	8	0	1	13
		cohérence	0	7	0	6	0	0	13
EXPRIMER	individuels	note	0	1	0	11	0	0	13
		étendue	0	3	0	9	0	1	13
		correction	0	5	0	8	0	0	13
		aisance	0	2	0	8	0	3	13
		interaction	0	1	0	10	0	2	13
		cohérence	0	1	0	8	0	4	13
	après discussion	note	0	0	0	13	0	0	13
		étendue	0	1	0	12	0	0	13
		correction	0	3	0	10	0	0	13
		aisance	0	3	0	9	0	1	13
		interaction	0	0	0	12	0	1	13
		cohérence	0	1	0	12	0	0	13
PRESENTER	individuels	note	0	0	0	11	0	2	13
		étendue	0	1	0	12	0	0	13
		correction	0	3	0	10	0	0	13
		aisance	0	1	0	11	0	1	13
		interaction	0	0	0	10	0	3	13
		cohérence	0	1	0	9	0	3	13
	après discussion	note	0	0	0	13	0	0	13
		étendue	0	0	0	13	0	0	13
		correction	0	3	0	10	0	0	13
		aisance	0	0	0	10	0	3	13
		interaction	0	0	0	10	0	3	13
		cohérence	0	0	0	10	0	3	13

3.2. Amina E.

3.2.1. Décrire

Avant et après la discussion, la majorité des experts, 9 sur 13, pensent qu'Amina E. a le niveau A1.1 à cette activité.

Lors du vote individuel, plus de la moitié des linguistes estiment qu'elle est de niveau A1 au critère *étendue*, et de niveau A1.1 aux 4 autres critères.

Après le travail de groupe, l'avis des experts concernant son niveau au critère *cohérence*, est partagé entre les niveaux A1.1-, A1.1 et A1. Plus de la moitié des linguistes estiment qu'elle a le niveau A1.1 aux 4 autres critères.

3.2.2. Exprimer

Lors du vote individuel, 7 experts donnent le niveau A1 à la candidate alors que 6 autres lui attribuent le niveau A1.1. L'avis des experts se partage donc entre ces deux niveaux. Le niveau d'Amina E. varie entre A1.1 et A1 suivant les critères observés. La majorité des experts estiment qu'elle a le niveau A1.1 aux critères *correction* et *cohérence* ; et le niveau A1 aux 3 autres critères.

Après discussion, une forte majorité, 10 experts, lui donne le niveau A1. L'avis des experts est partagé entre les niveaux A1.1 et A1 pour les critères *interaction* et *cohérence*. Pour les 3 autres critères, plus de la moitié des linguistes pensent qu'elle a le niveau A1.

3.2.3. Présenter

Avant et après discussion, la majorité des experts estiment que la candidate a un niveau global A1.1.

Lors des votes individuels, plus de la moitié des linguistes estiment qu'elle a le niveau A1.1 aux critères *étendue*, *aisance* et *cohérence* ; et le niveau A1 au critère *correction*. L'avis des experts oscille entre les niveaux A1.1 et A1 pour le critère *interaction*.

Après la discussion, ils lui donnent le niveau A1.1 aux critères *étendue*, *aisance* et *interaction*. Leurs votes se partagent entre les niveaux A1.1 et A1 pour la *correction* et la *cohérence*.

Niveaux du candidat par activité

Activité	Niveau
Décrire	A1.1
Exprimer	A1
Présenter	A1.1

Résultats des votes des experts

AMINA E	Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
DECRIRE	individuels	note	0	9	0	4	0	0	13
		étendue	0	5	0	8	0	0	13
		correction	0	8	0	5	0	0	13
		aisance	0	9	0	4	0	0	13
		interaction	2	7	0	4	0	0	13
		cohérence	1	9	0	3	0	0	13
	après discussion	note	0	9	0	4	0	0	13
		étendue	0	9	0	4	0	0	13
		correction	0	9	0	4	0	0	13
		aisance	0	9	0	4	0	0	13
		interaction	0	9	0	4	0	0	13
		cohérence	3	6	0	4	0	0	13
EXPRIMER	individuels	note	0	6	0	7	0	0	13
		étendue	0	5	0	8	0	0	13
		correction	0	8	1	4	0	0	13
		aisance	0	5	0	8	0	0	13
		interaction	2	3	0	8	0	0	13
		cohérence	1	7	0	5	0	0	13
	après discussion	note	0	3	0	10	0	0	13
		étendue	0	3	0	10	0	0	13
		correction	0	3	0	10	0	0	13
		aisance	0	3	0	10	0	0	13
		interaction	0	6	0	7	0	0	13
		cohérence	0	6	0	7	0	0	13
PRESENTER	individuels	note	0	8	0	4	0	1	13
		étendue	0	7	0	5	0	1	13
		correction	0	5	0	7	0	1	13
		aisance	0	8	0	4	0	1	13
		interaction	0	6	0	6	0	1	13
		cohérence	1	8	0	3	0	1	13
	après discussion	note	0	9	0	4	0	0	13
		étendue	0	9	0	4	0	0	13
		correction	0	6	0	7	0	0	13
		aisance	1	8	0	4	0	0	13
		interaction	1	8	0	4	0	0	13
		cohérence	3	6	0	4	0	0	13

3.3. Amina I.

3.3.1. Décrire

Avant et après le travail de groupe, la majorité des experts estiment qu'AMINA I. est de niveau A1 globalement et pour chaque critère.

3.3.2. Exprimer

Lors du vote individuel, l'opinion des experts est partagée : 7 d'entre eux pensent qu'elle a le niveau A1 et les 6 autres le niveau A1.1. Son niveau varie selon les critères. La majorité des experts lui donnent le niveau A1 aux critères : *interaction* et *cohérence* ; et le niveau A1.1 pour *l'aisance*. Pour les critères *correction* et *étendue*, leur avis est partagé entre les niveaux A1.1 et A1.

Après discussion, plus des trois quart des linguistes, 11 sur 13, pensent qu'elle est de niveau A1.1 globalement et pour chaque critère.

3.3.3. Présenter

Que ce soit avant ou après le travail de groupe, la plupart des linguistes pensent que la candidate est de niveau A1 globalement et pour chaque critère.

Niveaux du candidat par activité

Activité	Niveau
Décrire	A1
Exprimer	A1.1
Présenter	A1

Résultats des votes des experts

AMINA I.	Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
DECRIRE	individuels	note	0	0	0	12	0	1	13
		étendue	0	0	0	11	1	1	13
		correction	0	1	0	9	1	2	13
		aisance	0	0	0	9	1	3	13
		interaction	0	0	0	10	1	2	13
		cohérence	0	0	0	11	0	2	13
	après discussion	note	0	0	0	13	0	0	13
		étendue	0	0	0	13	0	0	13
		correction	0	0	0	13	0	0	13
		aisance	0	0	0	12	0	1	13
		interaction	0	0	0	12	0	1	13
		cohérence	0	0	0	12	0	1	13
EXPRIMER	individuels	note	0	6	0	7	0	0	13
		étendue	0	6	0	7	0	0	13
		correction	0	6	0	6	0	1	13
		aisance	0	8	0	5	0	0	13
		interaction	0	5	0	8	0	0	13
		cohérence	0	5	0	8	0	0	13
	après discussion	note	0	11	0	2	0	0	13
		étendue	0	11	0	2	0	0	13
		correction	0	11	0	2	0	0	13
		aisance	0	11	0	2	0	0	13
		interaction	0	11	0	2	0	0	13
		cohérence	0	11	0	2	0	0	13
PRESENTER	individuels	note	0	2	0	10	0	1	13
		étendue	0	3	0	9	0	1	13
		correction	0	2	0	10	0	1	13
		aisance	0	1	0	9	1	2	13
		interaction	0	1	0	10	1	1	13
		cohérence	0	4	0	7	1	1	13
	après discussion	note	0	0	0	13	0	0	13
		étendue	0	0	0	13	0	0	13
		correction	0	0	0	13	0	0	13
		aisance	0	0	0	13	0	0	13
		interaction	0	0	0	13	0	0	13
		cohérence	0	0	0	13	0	0	13

3.4. Belal

3.4.1. Décrire

Avant et après le travail de groupe, pratiquement tous les experts, 12 sur 13, estiment que ce candidat est de niveau A1. Pour les 5 critères, plus de la moitié des linguistes pensent qu'il est de ce niveau.

3.4.2. Exprimer

Que ce soit avant ou après la discussion, plus de la moitié des experts lui donnent un niveau A1 globalement et aux 5 critères.

3.4.3. Présenter

Avant et après le travail de groupe, la majorité des experts pensent qu'il est de niveau A1 globalement et aux 5 critères.

Niveaux du candidat par activité

Activité	Niveau
Décrire	A1
Exprimer	A1
Présenter	A1

Résultats des votes des experts

BELAL	Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
DECRIRE	individuels	note	0	1	0	12	0	0	13
		étendue	0	2	0	10	1	0	13
		correction	0	1	0	12	0	0	13
		aisance	0	1	0	9	2	1	13
		interaction	0	1	0	11	1	0	13
		cohérence	0	1	0	11	1	0	13
	après discussion	note	0	0	0	12	0	1	13
		étendue	0	0	0	12	0	1	13
		correction	0	3	0	9	0	1	13
		aisance	0	0	0	12	0	1	13
		interaction	0	3	0	9	0	1	13
		cohérence	0	0	0	12	0	1	13
EXPRIMER	individuels	note	0	2	0	10	0	0	13
		étendue	0	3	0	9	0	1	13
		correction	0	3	0	9	1	0	13
		aisance	0	2	0	7	1	3	13
		interaction	0	2	0	10	0	1	13
		cohérence	0	0	0	9	2	2	13
	après discussion	note	0	0	0	13	0	0	13
		étendue	0	3	0	10	0	0	13
		correction	0	3	0	10	0	0	13
		aisance	0	0	0	13	0	0	13
		interaction	0	0	0	13	0	0	13
		cohérence	0	0	0	13	0	0	13
PRESENTER	individuels	note	0	0	0	9	0	4	13
		étendue	0	0	0	8	2	3	13
		correction	0	0	0	9	2	2	13
		aisance	0	0	0	7	3	3	13
		interaction	0	0	0	7	4	2	13
		cohérence	0	0	0	7	1	5	13
	après discussion	note	0	0	0	13	0	0	13
		étendue	0	0	0	13	0	0	13
		correction	0	0	0	13	0	0	13
		aisance	0	0	0	13	0	0	13
		interaction	0	0	0	13	0	0	13
		cohérence	0	0	0	13	0	0	13

3.5. Felix

3.5.1. Décrire

Avant et après le travail de groupe, la plupart des experts estiment que Félix a le niveau A1.1- globalement et aux 5 critères.

3.5.2. Exprimer

Que ce soit avant ou après la discussion, plus des trois quart des experts lui donnent un niveau A1 globalement et aux 5 critères.

3.5.3. Présenter

Avant et après le travail de groupe, plus de la moitié des experts pensent que Félix a le niveau A1 globalement et aux 5 critères.

Niveaux du candidat par activité

Activité	Niveau
Décrire	A1.1-
Exprimer	A1.1-
Présenter	A1.1-

Résultats des votes des experts

FELIX	Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
DECRIRE	individuels	note	11	2	0	0	0	0	13
		étendue	7	5	0	1	0	0	13
		correction	11	2	0	0	0	0	13
		aisance	12	1	0	0	0	0	13
		interaction	12	1	0	0	0	0	13
		cohérence	11	2	0	0	0	0	13
	après discussion	note	13	0	0	0	0	0	13
		étendue	10	3	0	0	0	0	13
		correction	13	0	0	0	0	0	13
		aisance	13	0	0	0	0	0	13
		interaction	13	0	0	0	0	0	13
		cohérence	13	0	0	0	0	0	13
EXPRIMER	individuels	note	12	1	0	0	0	0	13
		étendue	10	3	0	0	0	0	13
		correction	11	2	0	0	0	0	13
		aisance	11	2	0	0	0	0	13
		interaction	12	1	0	0	0	0	13
		cohérence	12	1	0	0	0	0	13
	après discussion	note	13	0	0	0	0	0	13
		étendue	12	1	0	0	0	0	13
		correction	13	0	0	0	0	0	13
		aisance	13	0	0	0	0	0	13
		interaction	13	0	0	0	0	0	13
		cohérence	13	0	0	0	0	0	13
PRESENTER	individuels	note	9	4	0	0	0	0	13
		étendue	9	4	0	0	0	0	13
		correction	10	3	0	0	0	0	13
		aisance	11	2	0	0	0	0	13
		interaction	9	4	0	0	0	0	13
		cohérence	10	3	0	0	0	0	13
	après discussion	note	10	3	0	0	0	0	13
		étendue	10	3	0	0	0	0	13
		correction	10	3	0	0	0	0	13
		aisance	10	3	0	0	0	0	13
		interaction	10	3	0	0	0	0	13
		cohérence	10	3	0	0	0	0	13

3.6. Ganista

3.6.1. Décrire

Plus de la moitié des experts, que ce soit avant ou après discussion, pensent que Ganista a le niveau A1 globalement et à chaque critère.

3.6.2. Exprimer

Lors du vote individuel, l'opinion des experts est partagée : 7 d'entre eux estiment que Ganista a le niveau A1.1 et les 6 autres le niveau A1. Son niveau varie selon les critères. La majorité des experts lui donnent le niveau A1 aux critères *étendue* et *interaction* ; et le niveau A1.1 pour *l'aisance* et la *cohérence*. Pour le critère *correction*, leur avis est partagé entre les niveaux A1.1 et le A1.

Après discussion, la majorité des experts, 9 sur 13, pensent qu'elle a le niveau A1.1 globalement et à chaque critère.

3.6.3. Présenter

La majorité des experts, avant et après le travail de groupe, pensent que Ganista a le niveau A1 globalement et à chaque critère.

Niveaux du candidat par activité

Activité	Niveau
Décrire	A1
Exprimer	A1.1
Présenter	A1

Résultats des votes des experts

GANISTA	Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
DECRIRE	individuels	note	0	5	0	8	0	0	13
		étendue	0	5	0	8	0	0	13
		correction	0	3	0	10	0	0	13
		aisance	0	2	0	10	0	1	13
		interaction	0	5	0	7	0	1	13
		cohérence	0	5	0	8	0	0	13
	après discussion	note	0	0	0	13	0	0	13
		étendue	0	0	0	13	0	0	13
		correction	0	0	0	13	0	0	13
		aisance	0	0	0	13	0	0	13
		interaction	0	0	0	13	0	0	13
		cohérence	0	0	0	13	0	0	13
EXPRIMER	individuels	note	0	7	0	6	0	0	13
		étendue	0	4	1	8	0	0	13
		correction	0	6	1	6	0	0	13
		aisance	0	12	0	1	0	0	13
		interaction	0	5	0	8	0	0	13
		cohérence	0	10	0	3	0	0	13
	après discussion	note	0	9	0	4	0	0	13
		étendue	0	10	0	3	0	0	13
		correction	0	10	0	3	0	0	13
		aisance	0	10	0	3	0	0	13
		interaction	0	10	0	3	0	0	13
		cohérence	0	10	0	3	0	0	13
PRESENTER	individuels	note	0	3	0	9	0	1	13
		étendue	0	1	0	10	0	2	13
		correction	0	2	0	10	0	1	13
		aisance	0	1	0	10	1	1	13
		interaction	0	3	0	7	0	3	13
		cohérence	0	4	0	8	0	1	13
	après discussion	note	0	0	0	13	0	0	13
		étendue	0	0	0	13	0	0	13
		correction	0	0	0	13	0	0	13
		aisance	0	0	0	13	0	0	13
		interaction	0	0	0	13	0	0	13
		cohérence	0	0	0	13	0	0	13

3.7. Hala

3.7.1. Décrire

La plupart des linguistes estiment que Hala a le niveau A1.1 avant et après le travail de groupe.

Lors du vote individuel, son niveau varie selon les critères observés. La majorité des experts pensent que Hala a le niveau A1.1 pour la *correction* et la *cohérence*. Leur avis est partagé pour les critères : *étendue*, *aisance* et *interaction*.

Après discussion, tous les experts estiment que son niveau pour chaque critère est un niveau A1.1.

3.7.2. Exprimer

La majorité des experts, avant et après le travail de groupe, pensent que Hala est de niveau A1.1 globalement et pour chaque critère.

3.7.3. Présenter

Avant et après discussion, plus des trois quart des experts pensent que Hala est globalement de niveau A1.1.

Lors du vote individuel, son niveau varie selon les critères entre A1.1 et A1. Plus de la moitié des experts lui donnent le niveau A1.1 pour la *correction* et la *cohérence* ; et A1 pour l'*aisance*. Leur avis est partagé concernant l'*interaction* et l'*étendue*.

Après le travail de groupe, la plupart des expert estiment que Hala a le niveau A1.1 globalement et pour chaque critère.

Niveaux du candidat par activité

Activité	Niveau
Décrire	A1.1
Exprimer	A1.1
Présenter	A1.1

Résultats des votes des experts

HALA	Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
DECRIRE	individuels	note	0	10	0	3	0	0	13
		étendue	0	6	0	7	0	0	13
		correction	0	10	0	3	0	0	13
		aisance	0	6	0	7	0	0	13
		interaction	0	6	0	7	0	0	13
		cohérence	0	9	0	4	0	0	13
	après discussion	note	0	13	0	0	0	0	13
		étendue	0	13	0	0	0	0	13
		correction	0	13	0	0	0	0	13
		aisance	0	13	0	0	0	0	13
		interaction	0	13	0	0	0	0	13
		cohérence	0	13	0	0	0	0	13
EXPRIMER	individuels	note	0	13	0	0	0	0	13
		étendue	1	11	0	1	0	0	13
		correction	3	10	0	0	0	0	13
		aisance	0	11	0	2	0	0	13
		interaction	0	9	0	4	0	0	13
		cohérence	1	11	0	1	0	0	13
	après discussion	note	0	13	0	0	0	0	13
		étendue	0	13	0	0	0	0	13
		correction	0	13	0	0	0	0	13
		aisance	0	13	0	0	0	0	13
		interaction	0	13	0	0	0	0	13
		cohérence	0	13	0	0	0	0	13
PRESENTER	individuels	note	0	10	0	3	0	0	13
		étendue	0	6	0	7	0	0	13
		correction	0	9	0	4	0	0	13
		aisance	0	4	0	9	0	0	13
		interaction	0	7	0	6	0	0	13
		cohérence	0	10	0	3	0	0	13
	après discussion	note	0	13	0	0	0	0	13
		étendue	0	13	0	0	0	0	13
		correction	0	13	0	0	0	0	13
		aisance	0	10	0	3	0	0	13
		interaction	0	10	0	3	0	0	13
		cohérence	0	13	0	0	0	0	13

3.8. Mansour

3.8.1. Décrire

La majorité des experts pensent que le candidat a le niveau A1.1- avant et après discussion.

Lors du vote individuel, plus de la moitié des experts lui donnent le niveau A1.1 au critère *étendue* et le niveau A1.1- aux 4 autres critères.

Après discussion, plus des trois quart des linguistes estiment que le candidat a le niveau A1.1- globalement et pour chaque critère.

3.8.2. Exprimer

La majorité des experts pensent que le candidat a le niveau A1.1- avant et après discussion.

Lors du vote individuel, plus de la moitié des experts lui donnent le niveau A1.1- au critère *cohérence*. Pour les 4 autres critères, les avis sont assez partagés, son niveau oscille entre A1.1- et A1.1.

Après discussion, tous les experts estiment qu'il est de niveau A1.1- globalement et pour chaque critère.

3.8.3. Présenter

Avant et après le travail de groupe, plus des trois quarts des experts pensent que le candidat est de niveau A1.1 globalement.

Lors du vote individuel, plus des trois quart des linguistes estiment que le candidat a le niveau A1.1- pour chaque critère.

Après la discussion, leur avis est partagé concernant le critère *cohérence* entre les niveaux A1.1- et A1.1. Plus des trois quart des linguistes estiment que le candidat a le niveau A1.1- aux 4 autres critères.

Niveaux du candidat par activité

Activité	Niveau
Décrire	A1.1-
Exprimer	A1.1-
Présenter	A1.1

Résultats des votes des experts

MANSOUR	Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
DECRIRE	individuels	note	8	5	0	0	0	0	13
		étendue	4	8	0	1	0	0	13
		correction	8	5	0	0	0	0	13
		aisance	8	5	0	0	0	0	13
		interaction	8	5	0	0	0	0	13
		cohérence	8	4	0	1	0	0	13
	après discussion	note	13	0	0	0	0	0	13
		étendue	10	3	0	0	0	0	13
		correction	10	3	0	0	0	0	13
		aisance	13	0	0	0	0	0	13
		interaction	13	0	0	0	0	0	13
		cohérence	13	0	0	0	0	0	13
EXPRIMER	individuels	note	8	5	0	0	0	0	13
		étendue	6	7	0	0	0	0	13
		correction	6	7	0	0	0	0	13
		aisance	6	7	0	0	0	0	13
		interaction	7	6	0	0	0	0	13
		cohérence	9	4	0	0	0	0	13
	après discussion	note	13	0	0	0	0	0	13
		étendue	13	0	0	0	0	0	13
		correction	13	0	0	0	0	0	13
		aisance	13	0	0	0	0	0	13
		interaction	13	0	0	0	0	0	13
		cohérence	13	0	0	0	0	0	13
PRESENTER	individuels	note	1	12	0	0	0	0	13
		étendue	0	12	0	1	0	0	13
		correction	2	11	0	0	0	0	13
		aisance	1	12	0	0	0	0	13
		interaction	1	12	0	0	0	0	13
		cohérence	2	10	0	1	0	0	13
	après discussion	note	3	10	0	0	0	0	13
		étendue	0	13	0	0	0	0	13
		correction	3	10	0	0	0	0	13
		aisance	3	10	0	0	0	0	13
		interaction	3	10	0	0	0	0	13
		cohérence	6	7	0	0	0	0	13

3.9. Munazza

3.9.1. Décrire

Lors du vote individuel, la majorité des experts, 11 sur 13, donnent à la candidate le niveau A1.1. Après discussion, une partie d'entre eux a changé d'avis, plus de la moitié, 8 sur 13, pensent qu'elle est de niveau A1.1-.

Avant le travail de groupe, pour chaque critère, la plupart des experts estiment que la candidate est de niveau A1.1. Après discussion, plus de la moitié lui donnent le niveau A1.1- à chaque critère.

3.9.2. Exprimer

Lors du vote individuel, l'opinion des experts est partagée : 7 d'entre eux estiment que la candidate a le niveau A1.1 et les 6 autres le niveau A1.1-. Son niveau varie selon les critères. La majorité des experts lui donnent le niveau A1.1 aux critères *étendue* et *correction*, le niveau A1.1- pour *l'aisance*. Pour les critères *interaction* et *cohérence*, leur avis est partagé entre les niveaux A1.1- et le A1.1.

Après le travail de groupe, Il est toujours difficile de statuer sur le niveau de la candidate étant donné qu'une très faible majorité, 7 experts sur 13, lui attribue le niveau A1.1-. Cependant, son niveau semble plutôt se rapprocher de A1.1- car trois quarts des experts estiment que son niveau est A1.1- pour 2 critères : *interaction*, *cohérence* ; et 7 experts contre 6 pensent qu'elle est aussi de niveau A1.1- aux 3 critères restants : *étendue*, *correction*, *aisance*.

3.9.3. Présenter

Avant et après le travail de groupe, plus des trois quart des experts pensent qu'elle a le niveau A1.1.

Lors du vote individuel, plus de la moitié des linguistes lui donnent le niveau A1.1 aux critères *correction* et *cohérence* et le niveau A1 au critère *interaction*. Les avis sont partagés entre les niveaux A1.1 et A1 pour *l'aisance* et *l'étendue*.

Après le travail de groupe, plus des trois quart des experts estiment qu'elle a le niveau A1.1 à chaque critère.

Niveaux du candidat par activité

Activité	Niveau
Décrire	A1.1-
Exprimer	A1.1-
Présenter	A1.1

Résultats des votes des experts

MUNAZZA	Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux	
DECRIRE	individuels	note	2	11	0	0	0	0	13	
		étendue	2	11	0	0	0	0	13	
		correction	3	10	0	0	0	0	13	
		aisance	4	8	0	1	0	0	13	
		interaction	1	10	0	2	0	0	13	
		cohérence	3	10	0	0	0	0	13	
	après discussion	note	8	5	0	0	0	0	13	
		étendue	8	5	0	0	0	0	13	
		correction	8	5	0	0	0	0	13	
		aisance	8	5	0	0	0	0	13	
		interaction	8	5	0	0	0	0	13	
		cohérence	8	5	0	0	0	0	13	
	EXPRIMER	individuels	note	6	7	0	0	0	0	13
			étendue	4	8	0	1	0	0	13
correction			5	8	0	0	0	0	13	
aisance			8	5	0	0	0	0	13	
interaction			6	6	0	1	0	0	13	
cohérence			6	7	0	0	0	0	13	
après discussion		note	7	6	0	0	0	0	13	
		étendue	7	6	0	0	0	0	13	
		correction	7	6	0	0	0	0	13	
		aisance	7	6	0	0	0	0	13	
		interaction	10	3	0	0	0	0	13	
		cohérence	10	3	0	0	0	0	13	
PRESENTER		individuels	note	0	11	0	2	0	0	13
			étendue	0	7	0	6	0	0	13
	correction		0	9	0	4	0	0	13	
	aisance		0	6	0	7	0	0	13	
	interaction		0	5	0	8	0	0	13	
	cohérence		1	9	0	3	0	0	13	
	après discussion	note	0	13	0	0	0	0	13	
		étendue	0	13	0	0	0	0	13	
		correction	0	13	0	0	0	0	13	
		aisance	0	10	0	3	0	0	13	
		interaction	0	10	0	3	0	0	13	
		cohérence	0	13	0	0	0	0	13	

3.10. Salwa

3.10.1. Décrire

Que ce soit avant ou après le travail de groupe, tous les experts estiment que Salwa est de niveau A1.1 globalement et plus des trois quart pensent qu'elle a le même niveau aux 5 critères.

3.10.2. Exprimer

Avant ou après la discussion, pratiquement tous les experts pensent que la candidate a le niveau A1.1 globalement et aux 5 critères.

3.10.3. Présenter

La plupart des linguistes pensent qu'elle est de niveau A1.1 globalement et pour chaque critère, avant et après la discussion.

Niveaux du candidat par activité

Activité	Niveau
Décrire	A1.1
Exprimer	A1.1
Présenter	A1.1

Résultats des votes des experts

SALWA	Votes	note	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
DECRIRE	individuels	étendue	0	11	0	0	0	0	11
		correction	0	11	0	0	0	0	11
		aisance	2	9	0	0	0	0	11
		interaction	0	11	0	0	0	0	11
		cohérence	0	11	0	0	0	0	11
		note	2	9	0	0	0	0	11
	après discussion	étendue	0	11	0	0	0	0	11
		correction	0	11	0	0	0	0	11
		aisance	0	11	0	0	0	0	11
		interaction	0	11	0	0	0	0	11
		cohérence	0	11	0	0	0	0	11
		note	0	11	0	0	0	0	11
EXPRIMER	individuels	étendue	0	11	0	0	0	0	11
		correction	0	10	0	1	0	0	11
		aisance	1	9	0	1	0	0	11
		interaction	0	10	0	1	0	0	11
		cohérence	0	10	0	1	0	0	11
		note	1	10	0	0	0	0	11
	après discussion	étendue	0	11	0	0	0	0	11
		correction	0	11	0	0	0	0	11
		aisance	0	11	0	0	0	0	11
		interaction	0	11	0	0	0	0	11
		cohérence	0	11	0	0	0	0	11
		note	0	11	0	0	0	0	11
PRESENTER	individuels	étendue	1	10	0	0	0	0	11
		correction	0	11	0	0	0	0	11
		aisance	1	10	0	0	0	0	11
		interaction	3	8	0	0	0	0	11
		cohérence	2	9	0	0	0	0	11
		note	3	8	0	0	0	0	11
	après discussion	étendue	1	10	0	0	0	0	11
		correction	1	10	0	0	0	0	11
		aisance	1	10	0	0	0	0	11
		interaction	1	10	0	0	0	0	11
		cohérence	1	10	0	0	0	0	11
		note	1	10	0	0	0	0	11

3.11. Yazid

3.11.1. Décrire

Lors du vote individuel, la majorité des experts estiment que le niveau de Yazid est A1.1. Plus de la moitié lui donnent le niveau A1.1 aux critères *étendue* et *interaction*. Leur avis est partagé concernant les 3 autres les critères entre les niveaux A1.1- et A1.1.

Après le travail de groupe, l'opinion des experts change, plus de la moitié d'entre eux pensent qu'il a le niveau A1.1- globalement et à chaque critère.

3.11.2. Exprimer

Lors du vote individuel, l'avis des experts est partagé : 6 linguistes lui donnent le niveau A1.1- contre 5 le niveau A1.1. Plus de la moitié des experts pensent qu'il a le niveau A1.1- aux critères *correction*, *aisance* et *cohérence*. Leur opinion est partagée concernant les critères *étendue* et *interaction* entre les niveaux A1.1- et A1.1.

Après discussion, plus de la moitié des experts, 7 sur 11, estiment qu'il est de niveau A1.1-.

3.11.3. Présenter

Avant et après le travail de groupe, plus de la moitié des experts pensent que Yazid est de niveau A1.1 globalement et pour chaque critère.

Niveaux du candidat par activité

Activité	Niveau
Décrire	A1.1-
Exprimer	A1.1-
Présenter	A1.1

Résultats des votes des experts

YAZID	Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
DECRIRE	individuels	note	4	7	0	0	0	0	11
		étendue	3	8	0	0	0	0	11
		correction	6	5	0	0	0	0	11
		aisance	6	5	0	0	0	0	11
		interaction	3	7	0	1	0	0	11
		cohérence	5	6	0	0	0	0	11
	après discussion	note	7	4	0	0	0	0	11
		étendue	7	4	0	0	0	0	11
		correction	7	4	0	0	0	0	11
		aisance	7	4	0	0	0	0	11
		interaction	7	4	0	0	0	0	11
		cohérence	7	4	0	0	0	0	11
EXPRIMER	individuels	note	6	5	0	0	0	0	11
		étendue	6	5	0	0	0	0	11
		correction	7	4	0	0	0	0	11
		aisance	8	3	0	0	0	0	11
		interaction	5	6	0	0	0	0	11
		cohérence	8	3	0	0	0	0	11
	après discussion	note	7	4	0	0	0	0	11
		étendue	7	4	0	0	0	0	11
		correction	7	4	0	0	0	0	11
		aisance	7	4	0	0	0	0	11
		interaction	7	4	0	0	0	0	11
		cohérence	7	4	0	0	0	0	11
PRESENTER	individuels	note	3	8	0	0	0	0	11
		étendue	1	10	0	0	0	0	11
		correction	4	7	0	0	0	0	11
		aisance	3	7	0	1	0	0	11
		interaction	2	8	0	1	0	0	11
		cohérence	2	9	0	0	0	0	11
	après discussion	note	4	7	0	0	0	0	11
		étendue	0	11	0	0	0	0	11
		correction	4	7	0	0	0	0	11
		aisance	4	7	0	0	0	0	11
		interaction	4	7	0	0	0	0	11
		cohérence	4	7	0	0	0	0	11



Conclusion

Les votes des experts sont plus homogènes après la discussion en groupe de travail. Nous obtenons en nous basant sur les votes après discussion les niveaux suivants pour chaque candidat :

Niveau des candidats par activité

Candidats	Décrire	Exprimer	Présenter
AHMED	A1.1	A1	A1
AMINA E	A1.1	A1	A1.1
AMINA I.	A1	A1.1	A1
BELAL	A1	A1	A1
FELIX	A1.1-	A1.1-	A1.1-
GANISTA	A1	A1.1	A1
HALA	A1.1	A1.1	A1.1
MANSOUR	A1.1-	A1.1-	A1.1
MUNAZZA	A1.1-	A1.1-	A1.1
SALWA	A1.1	A1.1	A1.1
YAZID	A1.1-	A1.1-	A1.1



Annexes





Kappa pour plusieurs juges et plusieurs catégories (Fleiss)

Les calculs du coefficient kappa ont été effectués à l'aide du logiciel R, package *irr* (fonction *kappam.fleiss*).

Soit m juges qui doivent classer n observations dans k catégories ($m > 2$ et $k > 2$).

Le coefficient kappa peut être défini de la façon suivante :

$$K = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}$$

Pour la $j^{\text{ème}}$ catégorie :

m_{ij} : nombre de juges qui ont mis la $i^{\text{ème}}$ observation dans la $j^{\text{ème}}$ catégorie.

$$p_j = \frac{1}{nm} \sum_{i=1}^n m_{ij}$$

$$q_j = 1 - p_j$$

$$P_i = \frac{1}{m(m-1)} \sum_{j=1}^k m_{ij}(m_{ij} - 1)$$

Pourcentage de paires de juges en accord :

$$\bar{P} = \frac{1}{n} \sum_{i=1}^n P_i$$

Pourcentage de paires de juges jugeant de façon aléatoire

$$\bar{P}_e = \sum_{j=1}^k p_j q_j$$

Sous l'hypothèse des jugements aléatoire, nous retrouvons que $k \sim N(0, \bar{s}_0)$:

$$\bar{s}_0 = \frac{\sqrt{2}}{\sum_{j=1}^k p_j q_j \sqrt{nm(m-1)}} \sqrt{\left(\sum_{j=1}^k p_j q_j \right) - \sum_{j=1}^k p_j q_j (q_j - p_j)}$$

Kappa – DECRIRE – étendue

Coefficient k

	1	2	3	4	7	8	9	10	11	12	13	14	15
1	0	0,46	0,43	0,43	0	0,29	0,46	0,29	0,29	0,29	0,29	1	1
2	0,46	0	0,66	0,66	0,51	0,83	1	0,83	0,83	0,83	0,83	0,46	0,46
3	0,43	0,66	0	1	0,51	0,83	0,66	0,83	0,49	0,49	0,49	0,43	0,43
4	0,43	0,66	1	0	0,51	0,83	0,66	0,83	0,49	0,49	0,49	0,43	0,43
7	0	0,51	0,51	0,51	0	0,66	0,51	0,66	0,66	0,66	0,66	0	0
8	0,29	0,83	0,83	0,83	0,66	0	0,83	1	0,65	0,65	0,65	0,29	0,29
9	0,46	1	0,66	0,66	0,51	0,83	0	0,83	0,83	0,83	0,83	0,46	0,46
10	0,29	0,83	0,83	0,83	0,66	1	0,83	0	0,65	0,65	0,65	0,29	0,29
11	0,29	0,83	0,49	0,49	0,66	0,65	0,83	0,65	0	1	1	0,29	0,29
12	0,29	0,83	0,49	0,49	0,66	0,65	0,83	0,65	1	0	1	0,29	0,29
13	0,29	0,83	0,49	0,49	0,66	0,65	0,83	0,65	1	1	0	0,29	0,29
14	1	0,46	0,43	0,43	0	0,29	0,46	0,29	0,29	0,29	0,29	0	1
15	1	0,46	0,43	0,43	0	0,29	0,46	0,29	0,29	0,29	0,29	1	0

Statistique z

	1	2	3	4	7	8	9	10	11	12	13	14	15
1	0	1,83	1,64	1,64	0	1,16	1,83	1,16	1,16	1,16	1,16	3	3
2	1,83	0	2,8	2,8	2,36	3,51	4,24	3,51	3,51	3,51	3,51	1,83	1,83
3	1,64	2,8	0	4,16	2,32	3,48	2,8	3,48	2,04	2,04	2,04	1,64	1,64
4	1,64	2,8	4,16	0	2,32	3,48	2,8	3,48	2,04	2,04	2,04	1,64	1,64
7	0	2,36	2,32	2,32	0	2,97	2,36	2,97	2,97	2,97	2,97	0	0
8	1,16	3,51	3,48	3,48	2,97	0	3,51	4,16	2,72	2,72	2,72	1,16	1,16
9	1,83	4,24	2,8	2,8	2,36	3,51	0	3,51	3,51	3,51	3,51	1,83	1,83
10	1,16	3,51	3,48	3,48	2,97	4,16	3,51	0	2,72	2,72	2,72	1,16	1,16
11	1,16	3,51	2,04	2,04	2,97	2,72	3,51	2,72	0	4,16	4,16	1,16	1,16
12	1,16	3,51	2,04	2,04	2,97	2,72	3,51	2,72	4,16	0	4,16	1,16	1,16
13	1,16	3,51	2,04	2,04	2,97	2,72	3,51	2,72	4,16	4,16	0	1,16	1,16
14	3	1,83	1,64	1,64	0	1,16	1,83	1,16	1,16	1,16	1,16	0	3
15	3	1,83	1,64	1,64	0	1,16	1,83	1,16	1,16	1,16	1,16	3	0

p-value

	1	2	3	4	7	8	9	10	11	12	13	14	15
1	0	0,07	0,1	0,1	1	0,25	0,07	0,25	0,25	0,25	0,25	0	0
2	0,07	0	0,01	0,01	0,02	0	0	0	0	0	0	0,07	0,07
3	0,1	0,01	0	0	0,02	0	0,01	0	0,04	0,04	0,04	0,1	0,1
4	0,1	0,01	0	0	0,02	0	0,01	0	0,04	0,04	0,04	0,1	0,1
7	1	0,02	0,02	0,02	0	0	0,02	0	0	0	0	1	1
8	0,25	0	0	0	0	0	0	0	0,01	0,01	0,01	0,25	0,25
9	0,07	0	0,01	0,01	0,02	0	0	0	0	0	0	0,07	0,07
10	0,25	0	0	0	0	0	0	0	0,01	0,01	0,01	0,25	0,25
11	0,25	0	0,04	0,04	0	0,01	0	0,01	0	0	0	0,25	0,25
12	0,25	0	0,04	0,04	0	0,01	0	0,01	0	0	0	0,25	0,25
13	0,25	0	0,04	0,04	0	0,01	0	0,01	0	0	0	0,25	0,25
14	0	0,07	0,1	0,1	1	0,25	0,07	0,25	0,25	0,25	0,25	0	0
15	0	0,07	0,1	0,1	1	0,25	0,07	0,25	0,25	0,25	0,25	0	0

Kappa –PRESENTER – aisance

Coefficient k

	1	2	3	4	7	8	9	10	11	12	13	14	15
1	0	0,308	0,379	0,379	0,379	0,379	0,308	0,308	0,379	0,379	0,379	1	0,778
2	0,308	0	0,495	0,495	0,495	0,495	1	1	0,476	0,476	0,476	0,308	0,143
3	0,379	0,495	0	1	1	1	0,495	0,495	0,596	0,596	0,596	0,379	0,393
4	0,379	0,495	1	0	1	1	0,495	0,495	0,596	0,596	0,596	0,379	0,393
7	0,379	0,495	1	1	0	1	0,495	0,495	0,596	0,596	0,596	0,379	0,393
8	0,379	0,495	1	1	1	0	0,495	0,495	0,596	0,596	0,596	0,379	0,393
9	0,308	1	0,495	0,495	0,495	0,495	0	1	0,476	0,476	0,476	0,308	0,143
10	0,308	1	0,495	0,495	0,495	0,495	1	0	0,476	0,476	0,476	0,308	0,143
11	0,379	0,476	0,596	0,596	0,596	0,596	0,476	0,476	0	1	1	0,379	0,217
12	0,379	0,476	0,596	0,596	0,596	0,596	0,476	0,476	1	0	1	0,379	0,217
13	0,379	0,476	0,596	0,596	0,596	0,596	0,476	0,476	1	1	0	0,379	0,217
14	1	0,308	0,379	0,379	0,379	0,379	0,308	0,308	0,379	0,379	0,379	0	0,778
15	0,778	0,143	0,393	0,393	0,393	0,393	0,143	0,143	0,217	0,217	0,217	0,778	0

Statistique z

	1	2	3	4	7	8	9	10	11	12	13	14	15
1	0	1,382	1,46	1,46	1,46	1,46	1,382	1,382	1,314	1,314	1,314	3,889	3,13
2	1,382	0	2,22	2,22	2,22	2,22	4,906	4,906	2,026	2,026	2,026	1,382	0,652
3	1,46	2,22	0	3,819	3,819	3,819	2,22	2,22	2,057	2,057	2,057	1,46	1,591
4	1,46	2,22	3,819	0	3,819	3,819	2,22	2,22	2,057	2,057	2,057	1,46	1,591
7	1,46	2,22	3,819	3,819	0	3,819	2,22	2,22	2,057	2,057	2,057	1,46	1,591
8	1,46	2,22	3,819	3,819	3,819	0	2,22	2,22	2,057	2,057	2,057	1,46	1,591
9	1,382	4,906	2,22	2,22	2,22	2,22	0	4,906	2,026	2,026	2,026	1,382	0,652
10	1,382	4,906	2,22	2,22	2,22	2,22	4,906	0	2,026	2,026	2,026	1,382	0,652
11	1,314	2,026	2,057	2,057	2,057	2,057	2,026	2,026	0	3	3	1,314	0,83
12	1,314	2,026	2,057	2,057	2,057	2,057	2,026	2,026	3	0	3	1,314	0,83
13	1,314	2,026	2,057	2,057	2,057	2,057	2,026	2,026	3	3	0	1,314	0,83
14	3,889	1,382	1,46	1,46	1,46	1,46	1,382	1,382	1,314	1,314	1,314	0	3,13
15	3,13	0,652	1,591	1,591	1,591	1,591	0,652	0,652	0,83	0,83	0,83	3,13	0

p-value

	1	2	3	4	7	8	9	10	11	12	13	14	15
1	0	0,167	0,144	0,144	0,144	0,144	0,167	0,167	0,189	0,189	0,189	0	0,002
2	0,167	0	0,026	0,026	0,026	0,026	0	0	0,043	0,043	0,043	0,167	0,514
3	0,144	0,026	0	0	0	0	0,026	0,026	0,04	0,04	0,04	0,144	0,112
4	0,144	0,026	0	0	0	0	0,026	0,026	0,04	0,04	0,04	0,144	0,112
7	0,144	0,026	0	0	0	0	0,026	0,026	0,04	0,04	0,04	0,144	0,112
8	0,144	0,026	0	0	0	0	0,026	0,026	0,04	0,04	0,04	0,144	0,112
9	0,167	0	0,026	0,026	0,026	0,026	0	0	0,043	0,043	0,043	0,167	0,514
10	0,167	0	0,026	0,026	0,026	0,026	0	0	0,043	0,043	0,043	0,167	0,514
11	0,189	0,043	0,04	0,04	0,04	0,04	0,043	0,043	0	0,003	0,003	0,189	0,406
12	0,189	0,043	0,04	0,04	0,04	0,04	0,043	0,043	0,003	0	0,003	0,189	0,406
13	0,189	0,043	0,04	0,04	0,04	0,04	0,043	0,043	0,003	0,003	0	0,189	0,406
14	0	0,167	0,144	0,144	0,144	0,144	0,167	0,167	0,189	0,189	0,189	0	0,002
15	0,002	0,514	0,112	0,112	0,112	0,112	0,514	0,514	0,406	0,406	0,406	0,002	0

Kappa –PRESENTER – interaction

Coefficient k

	1	2	3	4	7	8	9	10	11	12	13	14	15
1	0	0,308	0,191	0,191	0,191	0,191	0,308	0,308	0,217	0,217	0,217	1	0,766
2	0,308	0	0,65	0,65	0,65	0,65	1	1	0,625	0,625	0,625	0,308	0,143
3	0,191	0,65	0	1	1	1	0,65	0,65	0,596	0,596	0,596	0,191	0,191
4	0,191	0,65	1	0	1	1	0,65	0,65	0,596	0,596	0,596	0,191	0,191
7	0,191	0,65	1	1	0	1	0,65	0,65	0,596	0,596	0,596	0,191	0,191
8	0,191	0,65	1	1	1	0	0,65	0,65	0,596	0,596	0,596	0,191	0,191
9	0,308	1	0,65	0,65	0,65	0,65	0	1	0,625	0,625	0,625	0,308	0,143
10	0,308	1	0,65	0,65	0,65	0,65	1	0	0,625	0,625	0,625	0,308	0,143
11	0,217	0,625	0,596	0,596	0,596	0,596	0,625	0,625	0	1	1	0,217	0,053
12	0,217	0,625	0,596	0,596	0,596	0,596	0,625	0,625	1	0	1	0,217	0,053
13	0,217	0,625	0,596	0,596	0,596	0,596	0,625	0,625	1	1	0	0,217	0,053
14	1	0,308	0,191	0,191	0,191	0,191	0,308	0,308	0,217	0,217	0,217	0	0,766
15	0,766	0,143	0,191	0,191	0,191	0,191	0,143	0,143	0,053	0,053	0,053	0,766	0

Statistique z

	1	2	3	4	7	8	9	10	11	12	13	14	15
1	0	1,382	0,773	0,773	0,773	0,773	1,382	1,382	0,83	0,83	0,83	3,889	2,696
2	1,382	0	2,771	2,771	2,771	2,771	4,597	4,597	2,454	2,454	2,454	1,382	0,652
3	0,773	2,771	0	3,819	3,819	3,819	2,771	2,771	2,057	2,057	2,057	0,773	0,773
4	0,773	2,771	3,819	0	3,819	3,819	2,771	2,771	2,057	2,057	2,057	0,773	0,773
7	0,773	2,771	3,819	3,819	0	3,819	2,771	2,771	2,057	2,057	2,057	0,773	0,773
8	0,773	2,771	3,819	3,819	3,819	0	2,771	2,771	2,057	2,057	2,057	0,773	0,773
9	1,382	4,597	2,771	2,771	2,771	2,771	0	4,597	2,454	2,454	2,454	1,382	0,652
10	1,382	4,597	2,771	2,771	2,771	2,771	4,597	0	2,454	2,454	2,454	1,382	0,652
11	0,83	2,454	2,057	2,057	2,057	2,057	2,454	2,454	0	3	3	0,83	0,213
12	0,83	2,454	2,057	2,057	2,057	2,057	2,454	2,454	3	0	3	0,83	0,213
13	0,83	2,454	2,057	2,057	2,057	2,057	2,454	2,454	3	3	0	0,83	0,213
14	3,889	1,382	0,773	0,773	0,773	0,773	1,382	1,382	0,83	0,83	0,83	0	2,696
15	2,696	0,652	0,773	0,773	0,773	0,773	0,652	0,652	0,213	0,213	0,213	2,696	0

p-value

	1	2	3	4	7	8	9	10	11	12	13	14	15
1	0	0,167	0,44	0,44	0,44	0,44	0,167	0,167	0,406	0,406	0,406	0	0,007
2	0,167	0	0,006	0,006	0,006	0,006	0	0	0,014	0,014	0,014	0,167	0,514
3	0,44	0,006	0	0	0	0	0,006	0,006	0,04	0,04	0,04	0,44	0,44
4	0,44	0,006	0	0	0	0	0,006	0,006	0,04	0,04	0,04	0,44	0,44
7	0,44	0,006	0	0	0	0	0,006	0,006	0,04	0,04	0,04	0,44	0,44
8	0,44	0,006	0	0	0	0	0,006	0,006	0,04	0,04	0,04	0,44	0,44
9	0,167	0	0,006	0,006	0,006	0,006	0	0	0,014	0,014	0,014	0,167	0,514
10	0,167	0	0,006	0,006	0,006	0,006	0	0	0,014	0,014	0,014	0,167	0,514
11	0,406	0,014	0,04	0,04	0,04	0,04	0,014	0,014	0	0,003	0,003	0,406	0,832
12	0,406	0,014	0,04	0,04	0,04	0,04	0,014	0,014	0,003	0	0,003	0,406	0,832
13	0,406	0,014	0,04	0,04	0,04	0,04	0,014	0,014	0,003	0,003	0	0,406	0,832
14	0	0,167	0,44	0,44	0,44	0,44	0,167	0,167	0,406	0,406	0,406	0	0,007
15	0,007	0,514	0,44	0,44	0,44	0,44	0,514	0,514	0,832	0,832	0,832	0,007	0



Calibrage des productions orales

Séminaire du 11 juillet 2006

Rapport d'analyse

Cellule qualité et expertises

Pôle évaluations et certifications



Sommaire

INTRODUCTION.....	5
1. LES DONNEES	7
1.1. LES PARTICIPANTS	7
1.2. LES PRODUCTIONS ORALES	7
2. EVALUATION DES EXPERTS.....	10
2.1. DEGRE D'ACCORD DES EXPERTS	10
2.2. ACCORD DES EXPERTS PAR NIVEAU	14
3. NIVEAUX DES CANDIDATS.....	17
3.1. AGNES.....	18
3.2. FATIMA.....	20
3.3. HAJRA.....	22
3.4. KEBIRA.....	24
3.5. MARGARET.....	26
3.6. VEERAPAN.....	28
CONCLUSION	31
ANNEXES	33
KAPPA POUR PLUSIEURS JUGES ET PLUSIEURS CATEGORIES (FLEISS).....	35



Introduction

L'objectif du séminaire organisé par le Centre International d'Etudes Pédagogiques et la Délégation générale à la langue française et aux langues de France, était de permettre aux experts réunis à cette occasion d'évaluer, selon les procédures définies par la Division des politiques linguistiques du Conseil de l'Europe, un certain nombre de productions orales de candidats au DILF (Diplôme initial de langue française), afin d'illustrer les différences existant entre le niveau A1.1 tel qu'il est défini dans le *Référentiel pour les premiers acquis en français* et le niveau A1 du *Cadre européen commun de référence pour les langues* (CECR).

Ces productions orales, après avoir été calibrées par la cellule qualité et expertise du CIEP, ont donné lieu à la création d'un DVD (*Productions orales et écrites illustrant, pour le français, les niveaux A1.1 et A1 du Cadre européen commun de référence pour les langues*).



1. Les données

1.1. Les participants

Les participants sont des experts en FLE (Français Langue Etrangère)¹, ils sont 8 au total (numérotés entre 1 et 8).

Ils estimeront tout d'abord individuellement le niveau des productions orales des 6 candidats présentés. Puis, dans un second temps, ils les analyseront en grand groupe afin d'établir un consensus sur le niveau de compétence des candidats.

1.2. Les productions orales

En amont du séminaire, le CIEP a filmé des locuteurs débutants interrogés à l'oral et à l'écrit afin de constituer un corpus de productions pour les niveaux A.1.1 et A.1.

Les locuteurs ont effectué 3 tâches :

- décrire un lieu (*décrire*),
- indiquer la nature d'un problème de santé (*exprimer*),
- se présenter (*présenter*).

Le jour du séminaire, 6 productions orales ont été présentées aux experts. 2 d'entre-elles ne seront pas présentées dans le DVD pour des raisons techniques.

Les experts donnent une note globale à chaque activité et une note aux 5 critères suivants :

- étendue,
- correction,
- aisance,
- interaction,
- cohérence.

¹ Liste des experts : Gilles BRETON, Jean-Claude BEACCO, Dorothee DUPLEIX, Maud LAUNAY, Sophie ETIENNE, Perrine MICHODET, Johanna PANTHIER, Marie-Françoise ROTTIER

Les experts doivent évaluer le niveau des candidats entre les niveaux A1.1 - et A2 :
A1.1-, A1.1, A1.1+, A1, A1+, A2.

Nous étudierons en premier lieu l'homogénéité de la correction des juges, puis le niveau des candidats. Nous ne calibrerons pas les productions orales grâce au modèle de Rasch à multiple facettes car nous ne disposons pas d'un nombre assez important d'experts.



2. Evaluation des experts

2.1. Degré d'accord des experts

Nous souhaitons savoir si les experts jugent de la même façon le niveau des candidats. Pour mesurer le degré d'accord des experts, nous utiliserons le coefficient kappa (Cf. annexes p. 35).

Les 8 experts ont classé les 6 candidats en fonction de leur niveau en français (le niveau en français ici est composé de 6 catégories). Le coefficient kappa k mesure le degré d'accord entre les experts. Le coefficient k est compris entre 0 et 1. En cas d'accord parfait entre les experts, $k=1$. Si les experts jugent le niveau des candidats d'une façon aléatoire, $k=0$. Plus le coefficient k est élevé plus le degré d'accord des experts sera important.

Nous calculons le coefficient de concordance (kappa) uniquement pour les candidats qui ont été notés par les 8 juges. Dans le tableau ci-dessous, la colonne "nb sujets" indique le nombre de candidats à partir duquel on a calculé le coefficient kappa. Nous vérifions l'hypothèse que les experts jugent de façon aléatoire ($H_0 : "k=0"$). Si le test est significatif (si $p < 5\%$), nous rejetons cette hypothèse avec un risque "de se tromper" de 5% (risque de 1^{ère} espèce $\alpha = 5\%$).

Nous ne calculons pas le coefficient kappa, quand nous constatons que tous les experts sont d'accord, il s'agit d'un accord parfait et $k=1$.

Valeurs du coefficient kappa

Activités	Critères	Votes individuels				Votes après discussion			
		kappa	z	p	nb sujets	kappa	z	p	nb sujets
DECRIRE	note	0.768	10	0	4	0.837	15.9	0	5
	étendue	Accord parfait			4	0.837	15.9	0	5
	correction	0.559	8.27	0	4	Accord parfait			4
	aisance	0.796	16.1	0	5	0.827	14.6	0	5
	Interaction	Accord parfait			3	Accord parfait			5
	cohérence	0.797	10	0	4	Accord parfait			4
EXPRIMER	note	0.88	15.3	0	6	Accord parfait			6
	étendue	0.833	17	0	6	Accord parfait			6
	correction	0.738	13.2	0	6	Accord parfait			6
	aisance	0.942	20.3	0	6	Accord parfait			6
	interaction	0.863	17.3	0	5	Accord parfait			6
	cohérence	0.694	13.6	0	6	Accord parfait			6
PRESENTER	note	0.65	8.97	0	6	Accord parfait			6
	étendue	0.789	14.3	0	6	Accord parfait			6
	correction	0.758	11.9	0	6	Accord parfait			6
	aisance	0.609	11	0	6	Accord parfait			6
	interaction	0.471	6.89	0	6	Accord parfait			6
	cohérence	0.828	11.4	0	6	Accord parfait			6

Dans l'ensemble, le jugement des experts est plus homogène après le travail de groupe. Ils sont tous d'accord pour les notes globales et les critères des deux activités : exprimer et présenter.

Pour l'activité « décrire », le degré d'accord entre les experts est plus fort après le travail de groupe pour la note globale, l'aisance, l'interaction et la cohérence. Pour le critère correction, l'accord des experts est parfait lors du vote individuel, il est mesuré à partir de 4 candidats. Après le travail de groupe, cet accord se révèle très fort ($k=0,837$), mais il n'est pas parfait. Cependant, nous ne pouvons pas considérer que le degré d'accord est meilleur lors du vote individuel car il est établi à partir de 4 candidats, tandis qu'après la discussion, cet indice est calculé sur 5 candidats. Si nous calculons le degré de concordance après la discussion à partir des 4 candidats utilisés pour le calcul sur les votes individuels, l'accord des experts après discussion est parfait lui aussi.

Il existe des désaccords uniquement pour le cinquième candidat « Hajra » avant et après discussion (Cf. tableaux ci-après). Pour les votes individuels, nous n'avons pas utilisé ce candidat dans le calcul du coefficient kappa car un expert n'a pas donné son avis sur son niveau. Nous jugeons qu'il est plus intéressant de se baser sur les votes après discussion pour établir le niveau des candidats concernant ce critère car nous disposons de plus d'informations.

Le test d'hypothèse des jugements aléatoires est rejeté ($p < 5\%$) pour chaque note globale et chaque critère observé. Les résultats du test indiquent que les experts n'évaluent pas de façon aléatoire les productions orales des candidats.

Niveaux évalués par les experts

Niveau	Correspondance
A1.1-	1
A1.1	2
A1.1+	3
A1	4
A1+	5
A2	6

DECRIRE - Critère correction - Votes individuels

Candidat	Linguistes							
	1	2	3	4	5	6	7	8
AGNES	4	4	4	4	4	4	4	4
FATIMA	1	1	1	1	1	1	1	1
HAJRA	2	2	4	0	2	2	1	2
KEBIRA	4	4	4	4	4	4	4	4
MARGARET	0	0	0	0	0	0	0	0
VEERAPAN	1	1	1	1	1	1	1	1

DECRIRE - Critère correction - Votes après discussion

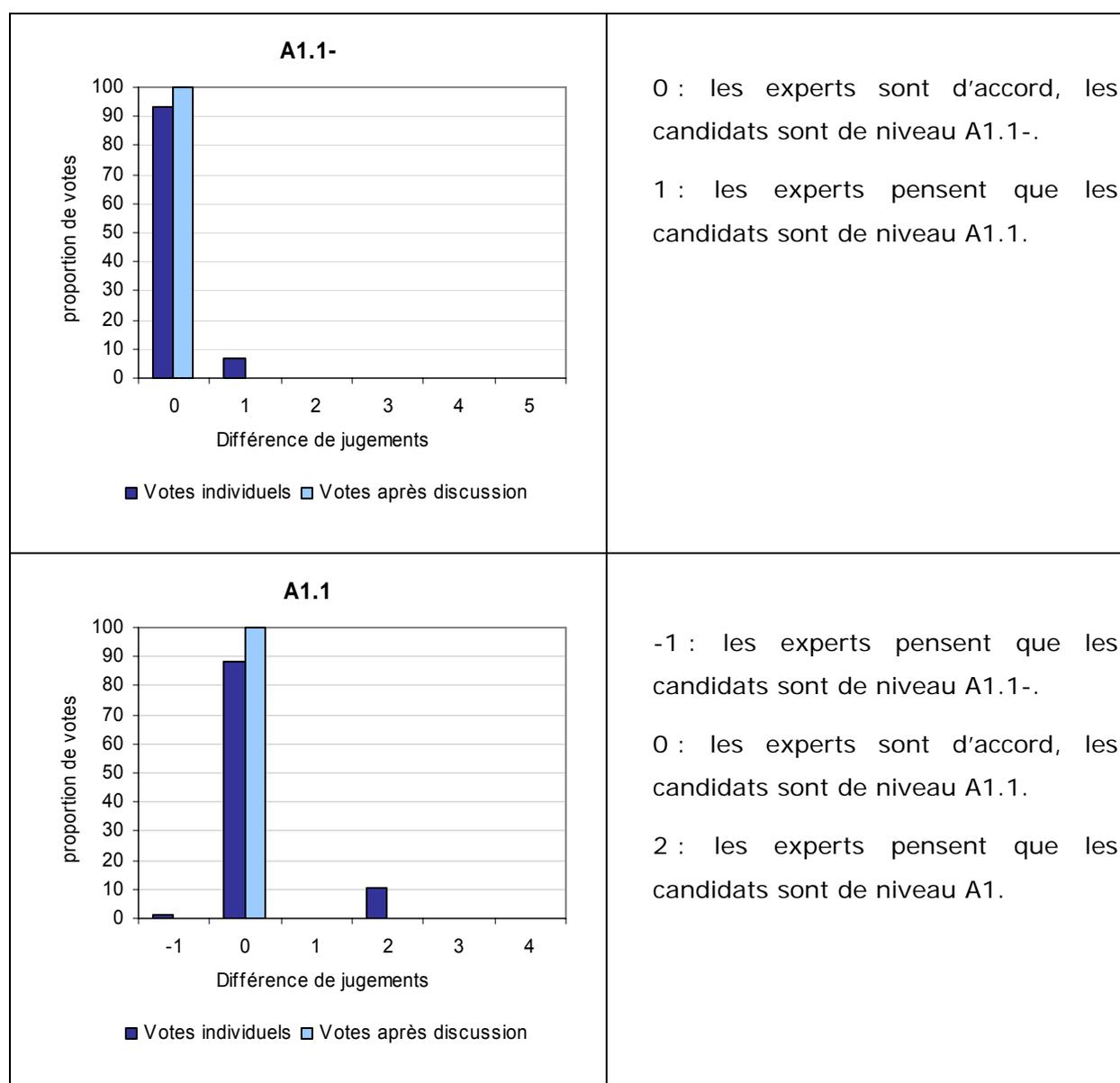
Candidat	Linguistes							
	1	2	3	4	5	6	7	8
AGNES	4	4	4	4	4	4	4	4
FATIMA	1	1	1	1	1	1	1	1
HAJRA	4	2	2	4	2	4	2	4
KEBIRA	5	5	5	5	5	5	5	5
MARGARET	0	0	0	0	0	0	0	0
VEERAPAN	1	1	1	1	1	1	1	1

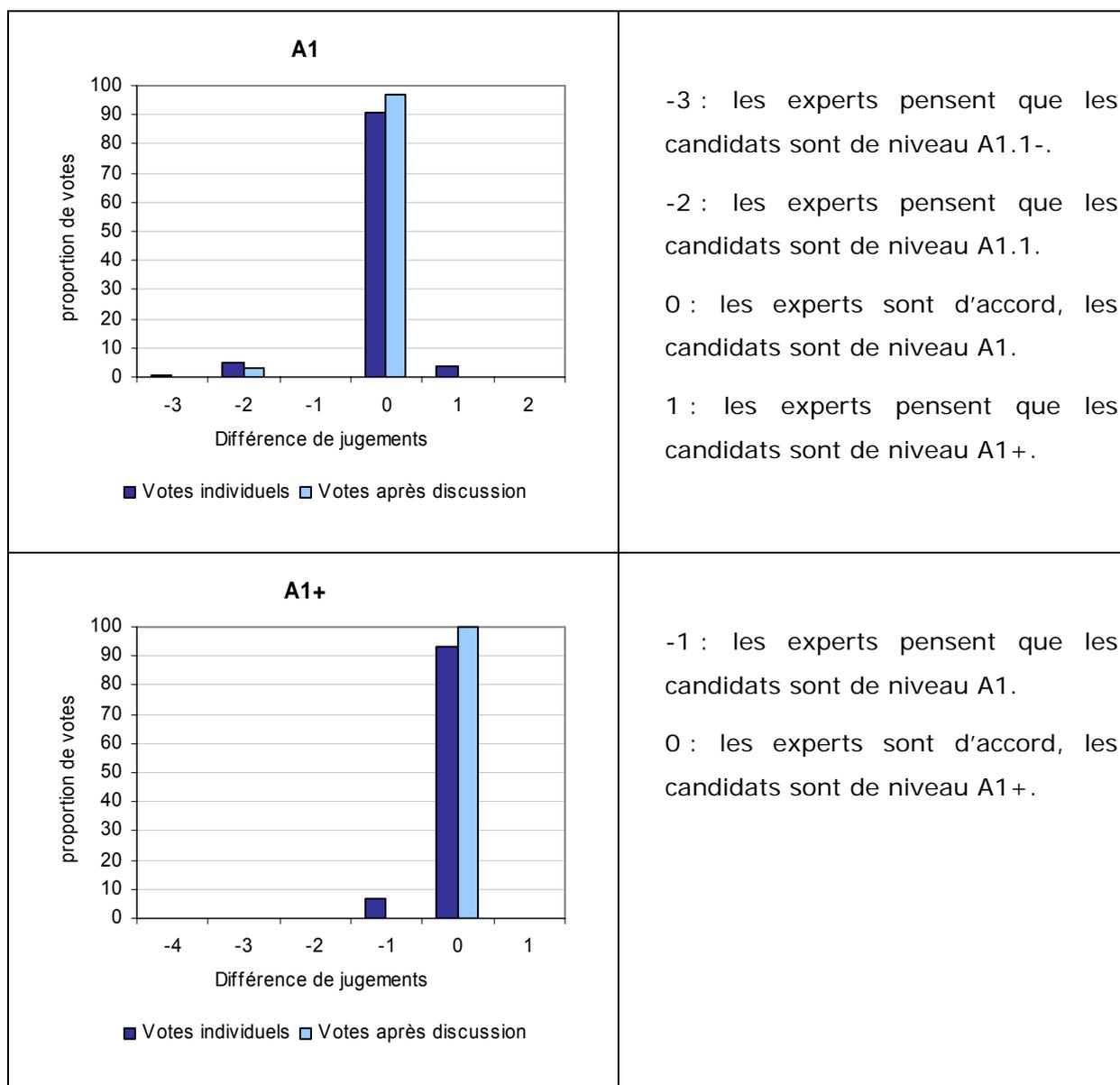
Le jugement des experts est plus homogène après le travail de groupe. Nous nous baserons ainsi sur les votes après discussion pour établir le niveau des candidats.

2.2. Accord des experts par niveau

Les graphiques suivants donnent une représentation de l'accord par niveau, avant et après le travail de groupe. Cette analyse est effectuée sur l'ensemble des notes globales et des critères. Cela nous permettra, par exemple, de savoir si les experts sont plus souvent en désaccord quand il s'agit d'attribuer le niveau A1.1- ou A1 à un candidat.

Pour chaque candidat, nous supposons que son niveau correspond au niveau le plus souvent sélectionné par les experts. Nous observons uniquement 4 niveaux possibles pour les candidats : A1.1-, A1.1, A1 et A1+.





Quel que soit le niveau de compétence des candidats, les experts sont plus souvent d'accord après le travail de groupe.

Lors du vote individuel, ils sont plus souvent d'accord quand les candidats sont estimés de niveaux A1.1- (93%), A1.1+ (93%) ou A1 (91%). Le pourcentage d'experts étant d'accord pour évaluer un candidat de niveau A1.1 est plus faible (88%).

Après discussion, Ils sont tous d'accord quand les candidats sont estimés de niveau A1.1- A1.1 ou A1.1+. Ils sont 97% à être d'accord quand le candidat est de niveau A1.

Les experts sont plus souvent d'accord sur le niveau des candidats après le travail de groupe.



3. Niveaux des candidats

Nous nous baserons sur les votes des experts après le travail de groupe pour évaluer le niveau des candidats. Pour chaque candidat, nous estimerons que son niveau correspond au niveau le plus souvent sélectionné par les experts.

3.1. Agnès

3.1.1. Décrire

Les experts n'ont pas changé pas d'avis après la discussion. Les 8 experts pensent qu'Agnès a un niveau global A1. Ils estiment qu'elle a aussi le niveau A1 aux critères *étendue*, *correction* et *cohérence* et le niveau A1+ aux deux autres critères.

3.1.2. Exprimer

L'opinion des experts ne change pas après le travail de groupe. Ils pensent toujours qu'Agnès a le niveau A1+ globalement et aux critères *étendue*, *aisance* et *interaction* ; et le niveau A1 aux deux autres critères.

3.1.3. Présenter

Après et avant discussion, la plupart des linguistes pensent qu'Agnès a un niveau global A1. Leur avis est partagé concernant les critères *étendue* et *correction* entre les niveaux A1 et A1.1. Plus de trois quart des experts estiment qu'elle a le niveau A1 aux critères *interaction* et *cohérence*.

3.1.4. Conclusion

Niveaux du candidat par activité

Activité	Niveau
Décrire	A1
Exprimer	A1+
Présenter	A1

Résultats des votes des experts

AGNES	Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
DECRIRE	Individuels	note	0	0	0	8	0	0	8
		étendue	0	0	0	8	0	0	8
		correction	0	0	0	8	0	0	8
		aisance	0	0	0	0	8	0	8
		interaction	0	0	0	0	8	0	8
		cohérence	0	0	0	8	0	0	8
	Après discussion	note	0	0	0	8	0	0	8
		étendue	0	0	0	8	0	0	8
		correction	0	0	0	8	0	0	8
		aisance	0	0	0	0	8	0	8
		interaction	0	0	0	0	8	0	8
		cohérence	0	0	0	8	0	0	8
EXPRIMER	Individuels	note	0	0	0	0	8	0	8
		étendue	0	0	0	0	8	0	8
		correction	0	0	0	8	0	0	8
		aisance	0	0	0	0	8	0	8
		interaction	0	0	0	0	8	0	8
		cohérence	0	0	0	8	0	0	8
	Après discussion	note	0	0	0	0	8	0	8
		étendue	0	0	0	0	8	0	8
		correction	0	0	0	8	0	0	8
		aisance	0	0	0	0	8	0	8
		interaction	0	0	0	0	8	0	8
		cohérence	0	0	0	8	0	0	8
PRESENTER	Individuels	note	0	0	0	7	1	0	8
		étendue	0	0	0	4	4	0	8
		correction	0	0	0	4	4	0	8
		aisance	0	0	0	5	3	0	8
		interaction	0	0	0	6	2	0	8
		cohérence	0	0	0	7	1	0	8
	Après discussion	note	0	0	0	8	0	0	8
		étendue	0	0	0	0	8	0	8
		correction	0	0	0	8	0	0	8
		aisance	0	0	0	8	0	0	8
		interaction	0	0	0	8	0	0	8
		cohérence	0	0	0	8	0	0	8

3.2. Fatima

3.2.1. Décrire

Lors du vote individuel, un peu plus de la moitié des experts, 5 sur 8, estiment que Fatima est de niveau A1.1- globalement. Son niveau varie selon les critères. La majorité des experts lui donnent le niveau A1.1 pour *l'aisance* et A1.1- pour *l'étendue*, la *correction* et la *cohérence*. Seulement 3 personnes sur 8 ont donné leur avis pour le critère *interaction*.

Après la discussion, tous les experts pensent qu'elle a le niveau A1.1- globalement et pour chaque critère.

Cette production ne sera pas présentée dans le DVD pour des raisons techniques.

3.2.2. Exprimer

Après et avant le travail de groupe, l'ensemble des experts pensent que Fatima est de niveau global A1.1. Tous les linguistes ou presque estiment qu'elle a le niveau A1.1- aux critères *étendue* et *correction* et le niveau A1 aux 3 autres critères.

3.2.3. Présenter

Que ce soit avant ou après la discussion, la plupart des experts pensent qu'elle est de niveau A1.1 globalement.

Lors du vote individuel, la majorité des experts lui donnent le niveau A1.1 aux 4 critères : *étendue*, *correction*, *aisance* et *cohérence*. Leur avis est partagé concernant *l'interaction* entre les niveaux A1.1 et A1.

Niveaux du candidat par activité

Activité	Niveau
Décrire	A1.1-
Exprimer	A1.1
Présenter	A1.1

Résultats des votes des experts

FATIMA	Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
DECRIRE	Individuels	note	5	3	0	0	0	0	8
		étendue	8	0	0	0	0	0	8
		correction	5	3	0	0	0	0	8
		aisance	1	7	0	0	0	0	8
		interaction	3	0	0	0	0	0	3
		cohérence	5	3	0	0	0	0	8
	Après discussion	note	8	0	0	0	0	0	8
		étendue	8	0	0	0	0	0	8
		correction	8	0	0	0	0	0	8
		aisance	8	0	0	0	0	0	8
		interaction	8	0	0	0	0	0	8
		cohérence	8	0	0	0	0	0	8
EXPRIMER	Individuels	note	0	8	0	0	0	0	8
		étendue	7	1	0	0	0	0	8
		correction	8	0	0	0	0	0	8
		aisance	0	8	0	0	0	0	8
		interaction	0	8	0	0	0	0	8
		cohérence	0	8	0	0	0	0	8
	Après discussion	note	0	8	0	0	0	0	8
		étendue	8	0	0	0	0	0	8
		correction	8	0	0	0	0	0	8
		aisance	0	8	0	0	0	0	8
		interaction	0	8	0	0	0	0	8
		cohérence	0	8	0	0	0	0	8
PRESENTER	Individuels	note	0	7	0	1	0	0	8
		étendue	0	7	0	1	0	0	8
		correction	0	8	0	0	0	0	8
		aisance	0	5	0	3	0	0	8
		interaction	0	4	0	4	0	0	8
		cohérence	0	8	0	0	0	0	8
	Après discussion	note	0	8	0	0	0	0	8
		étendue	0	8	0	0	0	0	8
		correction	0	8	0	0	0	0	8
		aisance	0	8	0	0	0	0	8
		interaction	0	8	0	0	0	0	8
		cohérence	0	8	0	0	0	0	8

3.3. Hajra

3.3.1. Décrire

Nous observons beaucoup de non réponses. Après le travail de groupe, seulement 4 experts ont donné un niveau à la candidate pour le critère *correction* et aucun ne s'est prononcé pour la *cohérence*. Pendant le vote individuel, un expert n'a pas indiqué le niveau global d'Hajra et aux critères *étendue*, *interaction* et *cohérence*.

Avant et après la discussion, les experts ont un avis très partagé concernant le niveau de la candidate, la moitié pensent qu'elle est de niveau A1.1 et l'autre de niveau A1.

Lors du vote individuel, la majorité des linguistes pensent qu'elle est de niveau A1.1 aux critères *étendue* et *correction* et de niveau A1 pour *l'aisance*, *l'interaction* et la *cohérence*.

Après le travail de groupe, les avis sont très partagés entre les niveaux A1.1 et A1 pour les critères *aisance* et *étendue*. L'ensemble des experts pensent qu'elle le niveau A1 pour *l'interaction*. Les experts n'étant pas arrivés à un consensus pour cette production, celle-ci ne sera pas présentée dans le DVD.

3.3.2. Exprimer

Après et avant le travail de groupe, l'ensemble des experts pense que Hajra est de niveau global A1.1.

Lors du vote individuel, l'avis des experts est partagé entre les niveaux A1.1 et A1 pour le critère *correction*. Tous les linguistes ou presque estiment qu'elle a le niveau A1 aux autres critères.

Après discussion, tous les experts lui ont attribué le niveau A1 à la note globale ainsi qu'aux 5 critères.

3.3.3. Présenter

Après et avant le travail de groupe, la majorité des experts pensent que Fatima est de niveau global A1.1.

Lors du vote individuel, la majorité des experts estiment qu'elle est de niveau A1.1 pour la correction et de niveau A1 pour les 4 autres critères.

Après discussion, tous les experts pensent qu'elle est de niveau A1 aux 5 critères.

Niveaux du candidat par activité

Activité	Niveau
Décrire	A1.1 et A1
Exprimer	A1
Présenter	A1

Résultats des votes des experts

HAJRA	Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
DECRIRE	Individuels	note	1	3	0	3	0	0	7
		étendue	1	5	0	1	0	0	7
		correction	1	5	0	2	0	0	8
		aisance	0	1	0	7	0	0	8
		interaction	0	2	0	5	0	0	7
		cohérence	1	0	0	6	0	0	7
	Après discussion	note	0	4	0	4	0	0	8
		étendue	0	4	0	4	0	0	8
		correction	0	4	0	0	0	0	4
		aisance	0	4	0	4	0	0	8
		interaction	0	0	0	8	0	0	8
		cohérence	0	0	0	0	0	0	0
EXPRIMER	Individuels	note	0	0	0	8	0	0	8
		étendue	0	0	0	8	0	0	8
		correction	0	4	0	4	0	0	8
		aisance	0	0	0	8	0	0	8
		interaction	0	1	0	7	0	0	8
		cohérence	0	1	0	7	0	0	8
	Après discussion	note	0	0	0	8	0	0	8
		étendue	0	0	0	8	0	0	8
		correction	0	0	0	8	0	0	8
		aisance	0	0	0	8	0	0	8
		interaction	0	0	0	8	0	0	8
		cohérence	0	0	0	8	0	0	8
PRESENTER	Individuels	note	0	3	0	5	0	0	8
		étendue	0	0	0	8	0	0	8
		correction	0	8	0	0	0	0	8
		aisance	0	2	0	6	0	0	8
		interaction	0	3	0	5	0	0	8
		cohérence	0	1	0	7	0	0	8
	Après discussion	note	0	0	0	8	0	0	8
		étendue	0	0	0	8	0	0	8
		correction	0	0	0	8	0	0	8
		aisance	0	0	0	8	0	0	8
		interaction	0	0	0	8	0	0	8
		cohérence	0	0	0	8	0	0	8

3.4. Kebira

3.4.1. Décrire

Lors du vote individuel, tous les experts donnent un niveau global A1 à Kebira. La plupart des experts estiment qu'elle a le niveau A1+ au critère *interaction* et le niveau A1 aux 4 autres.

Après discussion, les 8 experts changent d'avis, ils estiment que Kebira a le niveau A1+ globalement et aux critères *étendue*, *aisance* et *interaction*. Ils pensent qu'elle a le niveau A1 aux 2 autres critères.

3.4.2. Exprimer

Que ce soit avant ou après discussion, tous les linguistes estiment que la candidate a un niveau global A1. La plupart pense qu'elle le niveau A1 aux critères *étendue*, *correction* et *cohérence*, et le niveau A1.1 aux deux autres critères.

3.4.3. Présenter

Avant et après le travail de groupe, tous les experts estiment que Fatima a un niveau global A1. L'ensemble des experts pensent qu'elle a le niveau A1 aux critères *correction*, *interaction* et *cohérence*, et le niveau A1+ aux deux autres critères.

Les productions orales de Kebira ne seront pas présentées dans le DVD pour des raisons techniques (enregistrement difficilement audible).

Niveaux du candidat par activité

Activité	Niveau
Décrire	A1+
Exprimer	A1
Présenter	A1

Résultats des votes des experts

KEBIRA	Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
DECRIRE	Individuels	note	0	0	0	8	0	0	8
		étendue	0	0	0	8	0	0	8
		correction	0	0	0	8	0	0	8
		aisance	0	0	0	7	1	0	8
		interaction	0	0	0	0	8	0	8
		cohérence	0	0	0	8	0	0	8
	Après discussion	note	0	0	0	0	8	0	8
		étendue	0	0	0	0	8	0	8
		correction	0	0	0	8	0	0	8
		aisance	0	0	0	0	8	0	8
		interaction	0	0	0	0	8	0	8
		cohérence	0	0	0	8	0	0	8
EXPRIMER	Individuels	note	0	0	0	8	0	0	8
		étendue	0	0	0	8	0	0	8
		correction	0	0	0	8	0	0	8
		aisance	0	0	0	1	7	0	8
		interaction	0	0	0	1	7	0	8
		cohérence	0	0	0	5	3	0	8
	Après discussion	note	0	0	0	8	0	0	8
		étendue	0	0	0	8	0	0	8
		correction	0	0	0	8	0	0	8
		aisance	0	0	0	0	8	0	8
		interaction	0	0	0	0	8	0	8
		cohérence	0	0	0	8	0	0	8
PRESENTER	Individuels	note	0	0	0	8	0	0	8
		étendue	0	0	0	0	8	0	8
		correction	0	0	0	8	0	0	8
		aisance	0	0	0	0	8	0	8
		interaction	0	0	0	8	0	0	8
		cohérence	0	0	0	8	0	0	8
	Après discussion	note	0	0	0	8	0	0	8
		étendue	0	0	0	0	8	0	8
		correction	0	0	0	8	0	0	8
		aisance	0	0	0	0	8	0	8
		interaction	0	0	0	8	0	0	8
		cohérence	0	0	0	8	0	0	8

3.5. Margaret

3.5.1. Décrire

La prestation de Margaret n'a pas pu être évaluée car elle était inaudible.

3.5.2. Exprimer

Un expert ne s'est pas prononcé sur le niveau de la candidate pour le critère *interaction*.

Lors des votes individuels, trois quart des experts pensent qu'elle a le niveau A1.1 globalement et aux critères *correction*, *interaction* et *cohérence*, et le niveau A1 aux critères *étendue* et *aisance*.

Après le travail de groupe, tous les experts estiment qu'elle a le niveau A1 globalement et à chaque critère.

Cette production ne sera pas présentée dans le DVD pour des raisons techniques.

3.5.3. Présenter

Que ce soit avant ou après la discussion, l'ensemble des experts ou presque estiment que la candidate a un niveau A1 globalement et aux 5 critères.

Cette production ne sera pas présentée dans le DVD pour des raisons techniques.

Niveaux du candidat par activité

Activité	Niveau
Décrire	?
Exprimer	A1
Présenter	A1

Résultats des votes des experts

MARGARET	Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
DECRIRE	Individuels	note							
		étendue							
		correction							
		aisance							
		interaction							
		cohérence							
	Après discussion	note							
		étendue							
		correction							
		aisance							
		interaction							
		cohérence							
EXPRIMER	Individuels	i_note	0	6	0	2	0	0	8
		i_etendue	0	0	0	8	0	0	8
		i_correction	0	6	0	2	0	0	8
		i_aisance	0	0	0	8	0	0	8
		i_interaction	0	6	0	1	0	0	7
		i_cohérence	0	6	0	2	0	0	8
	Après discussion	g_note	0	0	0	8	0	0	8
		g_etendue	0	0	0	8	0	0	8
		g_correction	0	0	0	8	0	0	8
		g_aisance	0	0	0	8	0	0	8
		g_interaction	0	0	0	8	0	0	8
		g_cohérence	0	0	0	8	0	0	8
PRESENTER	Individuels	i_note	0	0	0	8	0	0	8
		i_etendue	0	0	0	8	0	0	8
		i_correction	0	1	0	7	0	0	8
		i_aisance	0	0	0	8	0	0	8
		i_interaction	0	0	0	8	0	0	8
		i_cohérence	0	0	0	8	0	0	8
	Après discussion	g_note	0	0	0	8	0	0	8
		g_etendue	0	0	0	8	0	0	8
		g_correction	0	0	0	8	0	0	8
		g_aisance	0	0	0	8	0	0	8
		g_interaction	0	0	0	8	0	0	8
		g_cohérence	0	0	0	8	0	0	8

3.6. Veerapan

3.6.1. Décrire

Que ce soit avant ou après le travail de groupe, les 8 experts estiment que Veerapan a le niveau A1.1- globalement et aux 5 critères.

3.6.2. Exprimer

Avant et après le travail de groupe, les 8 experts estiment que Veerapan a le niveau A1.1- globalement et aux 5 critères.

3.6.3. Présenter

Avant et après la discussion, l'ensemble des linguistes lui donnent le niveau A1.1 aux 5 critères et globalement.

Niveaux du candidat par activité

Activité	Niveau
Décrire	A1.1-
Exprimer	A1.1-
Présenter	A1.1

Résultats des votes des experts

VEERAPAN	Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
DECRIRE	Individuels	note	8	0	0	0	0	0	8
		étendue	8	0	0	0	0	0	8
		correction	8	0	0	0	0	0	8
		aisance	8	0	0	0	0	0	8
		interaction	8	0	0	0	0	0	8
		cohérence	8	0	0	0	0	0	8
	Après discussion	note	8	0	0	0	0	0	8
		étendue	8	0	0	0	0	0	8
		correction	8	0	0	0	0	0	8
		aisance	8	0	0	0	0	0	8
		interaction	8	0	0	0	0	0	8
		cohérence	8	0	0	0	0	0	8
EXPRIMER	Individuels	i_note	8	0	0	0	0	0	8
		i_etendue	8	0	0	0	0	0	8
		i_correction	8	0	0	0	0	0	8
		i_aisance	8	0	0	0	0	0	8
		i_interaction	8	0	0	0	0	0	8
		i_cohérence	8	0	0	0	0	0	8
	Après discussion	g_note	8	0	0	0	0	0	8
		g_etendue	8	0	0	0	0	0	8
		g_correction	8	0	0	0	0	0	8
		g_aisance	8	0	0	0	0	0	8
		g_interaction	8	0	0	0	0	0	8
		g_cohérence	8	0	0	0	0	0	8
PRESENTER	Individuels	i_note	0	8	0	0	0	0	8
		i_etendue	0	8	0	0	0	0	8
		i_correction	0	8	0	0	0	0	8
		i_aisance	0	8	0	0	0	0	8
		i_interaction	0	8	0	0	0	0	8
		i_cohérence	0	8	0	0	0	0	8
	Après discussion	g_note	0	8	0	0	0	0	8
		g_etendue	0	8	0	0	0	0	8
		g_correction	0	8	0	0	0	0	8
		g_aisance	0	8	0	0	0	0	8
		g_interaction	0	8	0	0	0	0	8
		g_cohérence	0	8	0	0	0	0	8



Conclusion

Les votes des experts sont plus homogènes après la discussion en groupe. Nous obtenons en nous basant sur les votes après discussion les niveaux suivants pour chaque candidat :

Niveau des candidats par activité

Candidats	Décrire	Exprimer	Présenter
Agnès	A1	A1+	A1
Fatima	A1.1-	A1.1	A1.1
Hajra	A1.1/A1	A1	A1
Kebira	A1+	A1	A1
Margaret	-	A1	A1
Veerapan	A1.1-	A1.1-	A1.1



Annexes





Kappa pour plusieurs juges et plusieurs catégories (Fleiss)

Les calculs du coefficient kappa ont été effectués à l'aide du logiciel R, package *irr* (fonction *kappam.fleiss*).

Soit m juges qui doivent classer n observations dans k catégories ($m > 2$ et $k > 2$).

Le coefficient kappa peut être défini de la façon suivante :

$$K = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}$$

Pour la $j^{\text{ème}}$ catégorie :

m_{ij} : nombre de juges qui ont mis la $i^{\text{ème}}$ observation dans la $j^{\text{ème}}$ catégorie.

$$p_j = \frac{1}{nm} \sum_{i=1}^n m_{ij}$$

$$q_j = 1 - p_j$$

$$P_i = \frac{1}{m(m-1)} \sum_{j=1}^k m_{ij}(m_{ij} - 1)$$

Pourcentage de paire de juges en accord :

$$\bar{P} = \frac{1}{n} \sum_{i=1}^n P_i$$

Pourcentage de pair de juges jugeant de façon aléatoire

$$\bar{P}_e = \sum_{j=1}^k p_j q_j$$

Sous l'hypothèse des jugements aléatoire, nous retrouvons que $k \sim N(0, \bar{s}_0)$:

$$\bar{s}_0 = \frac{\sqrt{2}}{\sum_{j=1}^k p_j q_j \sqrt{nm(m-1)}} \sqrt{\left(\sum_{j=1}^k p_j q_j \right) - \sum_{j=1}^k p_j q_j (q_j - p_j)}$$



Calibrage des productions écrites

Séminaire du 11 juillet

Rapport d'analyse

Cellule qualité et expertises

Pôle évaluations et certifications



Sommaire

INTRODUCTION.....	3
1. LES DONNEES	3
1.1. LES PARTICIPANTS	3
1.2. LES PRODUCTIONS ECRITES.....	3
2. EVALUATION DES EXPERTS.....	3
2.1. DEGRE D'ACCORD DES EXPERTS	3
2.2. ACCORD DES EXPERTS PAR NIVEAU.....	3
3. NIVEAUX DES PRODUCTIONS.....	3
3.1. COPIE 1	3
3.2. COPIE 2	3
3.3. COPIE 3	3
3.4. COPIE 4	3
3.5. COPIE 5	3
3.6. COPIE 6	3
3.7. COPIE 7	3
3.8. COPIE 8	3
3.9. COPIE 9	3
CONCLUSION	3
ANNEXES	3
KAPPA POUR PLUSIEURS JUGES ET PLUSIEURS CATEGORIES (FLEISS).....	3



Introduction

Le séminaire organisé le 11 juillet 2006 par le Centre International d'Etudes Pédagogiques et la Délégation générale à la langue française et aux langues de France, a eu pour objectif de sélectionner, selon les procédures définies par la Division des politiques linguistiques du Conseil de l'Europe, un certain nombre de productions écrites de candidats au DILF (Diplôme initial de langue française).

L'évaluation de ces productions écrites permet d'illustrer les différences existant entre le niveau A1.1 tel qu'il est défini dans le « *Référentiel pour les premiers acquis en français* » et le niveau A1 du *Cadre européen commun de référence pour les langues* (CECR).

Ces productions écrites, après avoir été calibrées par la cellule qualité et expertise du CIEP ont donné lieu à la création d'un DVD (*Productions orales et écrites illustrant, pour le français, les niveaux A1.1 et A1 du Cadre européen commun de référence pour les langues*).



1. Les données

1.1. Les experts

Les participants sont des experts¹ en Français langue étrangère, ils sont 8 au total (numérotés entre 1 et 8) et ont évalué 13 productions écrites. Cependant, après évaluation, 4 productions n'ont pas été retenues pour illustrer les niveaux A1.1 et A1 car elles présentaient certains problèmes (non respect de la consigne, manque de matière à évaluer) ou parce qu'elles n'étaient pas représentatives du public cible. L'analyse des niveaux portera donc sur 9 productions (copies numérotées de 1 à 9).

Dans un premier temps, les experts ont estimé individuellement le niveau de 9 productions écrites. Dans un second temps, une analyse, en grand groupe, a permis d'obtenir un consensus sur un niveau de compétence précis.

1.2. Les productions écrites

Les experts ont travaillé sur la tâche de production écrite :

- Laisser un message simple.

Les experts donnent une note globale à la production et une note par critère, soit 4 au total :

- correction sociolinguistique (*sociolinguistique*),
- étendue du vocabulaire (*étendue*),
- correction grammaticale et orthographe (*grammaire*),
- cohérence (*cohérence*).

Les experts doivent évaluer des productions en attribuant une note située entre les niveaux A1.1 - et A2, c'est-à-dire les niveaux : A1.1-, A1.1, A1.1+, A1, A1+, A2.

Les experts doivent également se prononcer sur le critère *adéquation*, en indiquant si oui ou non le candidat a bien respecté la consigne de l'exercice.

¹ Liste des experts : Gilles BRETON, Jean-Claude BEACCO, Dorothée DUPLEIX, Maud LAUNAY, Sophie ETIENNE, Perrine MICHODET, Johanna PANTHIER, Marie-Françoise ROTTIER

La note globale n'est pas obtenue en faisant la moyenne de ces 5 critères. Les critères *sociolinguistique, étendue, grammaire, cohérence* étayent le niveau global de la production et permettent d'affiner le profil linguistique du candidat à l'écrit.

Nous étudierons en premier lieu l'homogénéité de la correction des juges, puis le niveau des productions. Nous ne calibrerons pas les productions écrites grâce au modèle de Rasch à multiple facettes car nous ne disposons pas d'un nombre assez important d'experts.

2. Evaluation des experts

2.1. Degré d'accord des experts

Nous souhaitons savoir si les experts jugent de la même façon le niveau des productions. Nous utilisons pour mesurer le degré d'accord des experts le coefficient kappa (Cf. annexes p. 3). Pour cette analyse, le jugement sur les 13 productions a pu être établi car il s'agissait d'analyser le degré d'accord des experts sur l'ensemble de leurs évaluations.

Les 8 experts ont classé les 13 productions selon le niveau en français (le niveau en français ici est composé de 6 catégories). Le coefficient kappa k mesure le degré d'accord entre les experts. Le coefficient k est compris entre 0 et 1. En cas d'accord parfait entre les experts $k=1$. Si les experts jugent le niveau des productions d'une façon aléatoire, $k=0$. Plus le coefficient k est élevé plus le degré d'accord des experts sera important.

Nous calculons le coefficient de concordance (kappa) uniquement pour les candidats qui ont été notés par les 8 juges. Dans le tableau ci-dessous, la colonne "nb sujets" indique le nombre de productions à partir desquels on a calculé le coefficient kappa. Nous vérifions l'hypothèse que les experts jugent de façon aléatoire ($H_0 : "k=0"$). Si le test est significatif (si $p < 5\%$), nous rejetons cette hypothèse avec un risque "de se tromper" de 5% (risque de 1^{ère} espèce $\alpha = 5\%$).

Nous ne calculons pas le coefficient kappa, quand nous constatons que tous les experts sont d'accord, il s'agit d'un accord parfait ($k=1$).

Valeurs du coefficient kappa

Critères	Votes individuels				Votes après discussion			
	kappa	z	p	nb sujets	kappa	z	p	nb sujets
note	0.771	20.6	0	13	Accord parfait			13
sociolinguistique	0.437	13.4	0	13	Accord parfait			13
étendue	0.577	14.2	0	11	Accord parfait			13
grammaire	0.667	16.9	0	12	Accord parfait			13
cohérence	0.563	13.2	0	10	Accord parfait			12
adéquation	Accord parfait			13	Accord parfait			13

Avant la discussion, les experts sont tous d'accord quand il s'agit de noter l'*adéquation*. Le degré de concordance des experts est très bon pour la note globale et le critère *grammaire* (le coefficient kappa est proche de 1). Il est assez bon pour les critères *étendue* et *cohérence*. En revanche, le degré d'accord est plus faible pour le critère *sociolinguistique*. Dans l'ensemble, le jugement des experts est assez homogène sauf pour le critère *sociolinguistique*.

Après le travail de groupe, quels que soient les critères ou la note globale observés, les experts sont d'accord à l'unanimité.

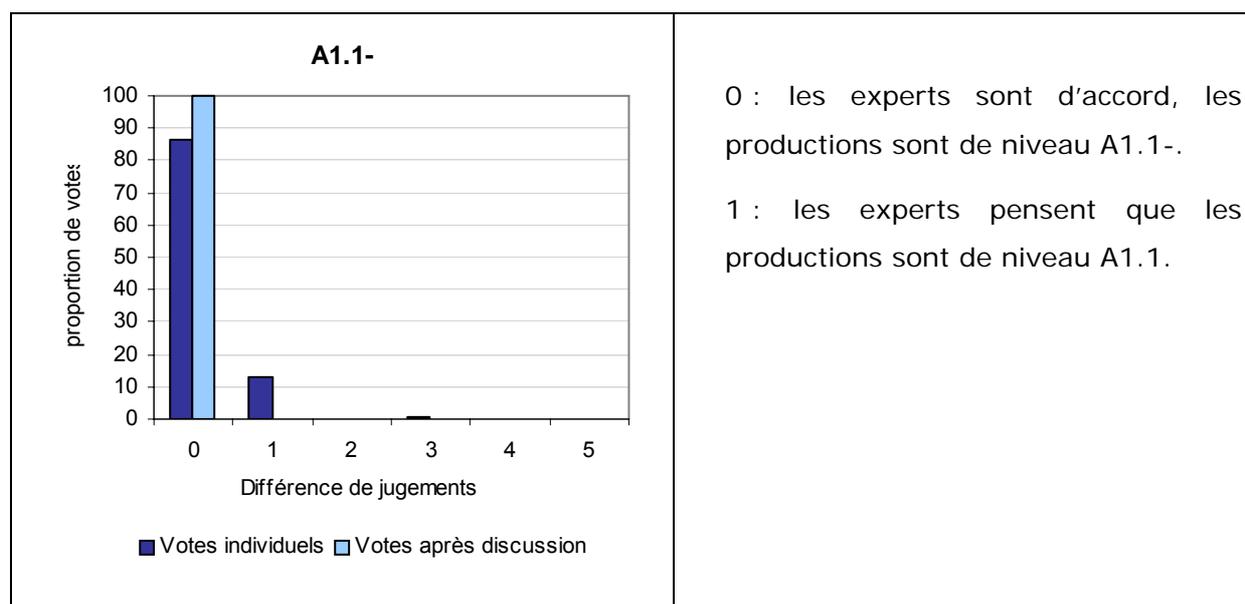
Le test d'hypothèse des jugements aléatoires est rejeté ($p < 5\%$) pour le niveau global et les 5 critères évalués avant et après le travail de groupe. Les résultats du test indiquent que les experts n'évaluent pas de façon aléatoire les productions écrites.

Le jugement des experts est plus homogène après le travail de groupe. Nous nous baserons ainsi sur les votes après discussion pour établir le niveau des productions.

2.2. Accord des experts par niveau

Les graphiques suivants donnent une représentation de l'accord par niveau, avant et après le travail de groupe. Cette analyse est effectuée sur l'ensemble des notes globales et des critères. Cela nous permettra, par exemple, de savoir si les experts sont plus souvent en désaccord quand il s'agit d'évaluer la compétence d'un candidat de niveau A1.1- ou celle d'un candidat de niveau A1.

Pour chaque production, nous supposons que son niveau correspond au niveau le plus souvent sélectionné par les experts. Nous observons uniquement 4 niveaux possibles pour les productions : A1.1-, A1.1, A1 et A1+.



<p style="text-align: center;">A1.1</p> <table border="1"> <caption>Data for A1.1</caption> <thead> <tr> <th>Différence de jugements</th> <th>Votes individuels (%)</th> <th>Votes après discussion (%)</th> </tr> </thead> <tbody> <tr> <td>-1</td> <td>15</td> <td>0</td> </tr> <tr> <td>0</td> <td>68</td> <td>100</td> </tr> <tr> <td>1</td> <td>0</td> <td>0</td> </tr> <tr> <td>2</td> <td>18</td> <td>0</td> </tr> <tr> <td>3</td> <td>0</td> <td>0</td> </tr> <tr> <td>4</td> <td>0</td> <td>0</td> </tr> </tbody> </table>	Différence de jugements	Votes individuels (%)	Votes après discussion (%)	-1	15	0	0	68	100	1	0	0	2	18	0	3	0	0	4	0	0	<p>-1 : les experts pensent que les productions sont de niveau A1.1-.</p> <p>0 : les experts sont d'accord, les productions sont de niveau A1.1.</p> <p>2 : les experts pensent que les productions sont de niveau A1.</p>
Différence de jugements	Votes individuels (%)	Votes après discussion (%)																				
-1	15	0																				
0	68	100																				
1	0	0																				
2	18	0																				
3	0	0																				
4	0	0																				
<p style="text-align: center;">A1.1+</p> <table border="1"> <caption>Data for A1.1+</caption> <thead> <tr> <th>Différence de jugements</th> <th>Votes individuels (%)</th> <th>Votes après discussion (%)</th> </tr> </thead> <tbody> <tr> <td>-2</td> <td>0</td> <td>0</td> </tr> <tr> <td>-1</td> <td>0</td> <td>0</td> </tr> <tr> <td>0</td> <td>88</td> <td>100</td> </tr> <tr> <td>1</td> <td>12</td> <td>0</td> </tr> <tr> <td>2</td> <td>0</td> <td>0</td> </tr> <tr> <td>3</td> <td>0</td> <td>0</td> </tr> </tbody> </table>	Différence de jugements	Votes individuels (%)	Votes après discussion (%)	-2	0	0	-1	0	0	0	88	100	1	12	0	2	0	0	3	0	0	<p>0 : les experts sont d'accord, les productions sont de niveau A1.1+.</p> <p>1 : les experts pensent que les productions sont de niveau A1.</p>
Différence de jugements	Votes individuels (%)	Votes après discussion (%)																				
-2	0	0																				
-1	0	0																				
0	88	100																				
1	12	0																				
2	0	0																				
3	0	0																				
<p style="text-align: center;">A1</p> <table border="1"> <caption>Data for A1</caption> <thead> <tr> <th>Différence de jugements</th> <th>Votes individuels (%)</th> <th>Votes après discussion (%)</th> </tr> </thead> <tbody> <tr> <td>-3</td> <td>2</td> <td>0</td> </tr> <tr> <td>-2</td> <td>5</td> <td>0</td> </tr> <tr> <td>-1</td> <td>0</td> <td>0</td> </tr> <tr> <td>0</td> <td>90</td> <td>100</td> </tr> <tr> <td>1</td> <td>2</td> <td>0</td> </tr> <tr> <td>2</td> <td>0</td> <td>0</td> </tr> </tbody> </table>	Différence de jugements	Votes individuels (%)	Votes après discussion (%)	-3	2	0	-2	5	0	-1	0	0	0	90	100	1	2	0	2	0	0	<p>-3 : les experts pensent que les productions sont de niveau A1.1-.</p> <p>-2 : les experts pensent que les productions sont de niveau A1.1.</p> <p>0 : les experts sont d'accord, les productions sont de niveau A1.</p> <p>1 : les experts pensent que les productions sont de niveau A1+.</p>
Différence de jugements	Votes individuels (%)	Votes après discussion (%)																				
-3	2	0																				
-2	5	0																				
-1	0	0																				
0	90	100																				
1	2	0																				
2	0	0																				

Lors du vote individuel, les linguistes sont plus souvent d'accord quand les productions sont estimées de niveaux A1 (90%) et A1.1- (87%), ils le sont un peu moins pour les productions de niveaux A1.1+ (80%). Le pourcentage d'experts étant d'accord pour évaluer une production de niveau A1.1 est plus faible (69%).

Après discussion, Ils sont tous d'accord quels que soient les niveaux des productions.

Les experts sont plus souvent d'accord sur le niveau des productions après le travail de groupe.

3. Niveaux des productions

Nous nous baserons sur les votes des experts après le travail de groupe pour évaluer le niveau des productions. Pour chacune d'entre elles, nous estimerons que son niveau correspond au niveau le plus souvent sélectionné par les experts.

3.1. Copie 1

Avant et après le travail de groupe, l'ensemble des experts ont attribué à cette production le niveau A1.1- globalement, ainsi qu'aux critères *cohérence* et *grammaire*. Plus de la moitié des experts donnent le niveau A1.1 au critère *sociolinguistique* et le niveau A1 au critère *étendue*.

Avant et après la discussion, tous les experts pensent que le candidat a respecté la consigne de l'exercice (critère *adéquation*).

Résultats des votes des experts

Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
Individuels	note	8	0	0	0	0	0	8
	sociolinguistique	1	7	0	0	0	0	8
	étendue	3	0	0	5	0	0	8
	grammaire	8	0	0	0	0	0	8
	cohérence	8	0	0	0	0	0	8
Après discussion	note	8	0	0	0	0	0	8
	sociolinguistique	0	8	0	0	0	0	8
	étendue	0	0	0	8	0	0	8
	grammaire	8	0	0	0	0	0	8
	cohérence	8	0	0	0	0	0	8

Résultats des votes des experts au critère adéquation

Votes	Critère	Oui	Non	Totaux
individuels	adéquation	8	0	8
après discussion	adéquation	8	0	8

Cette production est de niveau A1.1-.

3.2. Copie 2

Avant et après la discussion, tous les experts pensent que cette production est de niveau A1.1-. La majorité des experts donnent le niveau A1.1- aux critères *cohérence*, *sociolinguistique* et *grammaire* ; et le niveau A1.1 au critère *étendue*.

Avant et après la discussion, tous les experts pensent que le candidat n'a pas respecté la consigne de l'exercice (critère *adéquation*).

Résultats des votes des experts

Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
Individuels	note	8	0	0	0	0	0	8
	sociolinguistique	5	3	0	0	0	0	8
	étendue	0	7	0	1	0	0	8
	grammaire	8	0	0	0	0	0	8
	cohérence	6	2	0	0	0	0	8
Après discussion	note	8	0	0	0	0	0	8
	sociolinguistique	8	0	0	0	0	0	8
	étendue	0	8	0	0	0	0	8
	grammaire	8	0	0	0	0	0	8
	cohérence	8	0	0	0	0	0	8

Résultats des votes des experts au critère adéquation

Votes	Critère	Oui	Non	Totaux
individuels	adéquation	0	8	8
après discussion	adéquation	0	8	8

Cette production est de niveau A1.1-.

3.3. Copie 3

Que ce soit avant ou après la discussion, tous les experts pensent que la production est de niveau A1.1-.

Lors du vote individuel, au moins trois quart des experts donnent un niveau A1.1- aux critères *cohérence*, *sociolinguistique* et *grammaire*. Leur avis partagé entre les niveaux A1.1-, A1.1 et A1 pour le critère *étendue*.

Après le travail de groupe, tous les experts estiment que la production est de niveau A1.1- globalement ainsi qu'aux critères : *cohérence*, *sociolinguistique* et *grammaire*. Les 8 linguistes accordent le niveau A1.1 au critère *étendue*.

Avant et après la discussion, tous les experts pensent que le candidat n'a pas respecté la consigne de l'exercice (critère *adéquation*).

Résultats des votes des experts

Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
Individuels	note	8	0	0	0	0	0	8
	sociolinguistique	8	0	0	0	0	0	8
	étendue	2	3	0	3	0	0	8
	grammaire	8	0	0	0	0	0	8
	cohérence	6	0	0	0	0	0	6
Après discussion	note	8	0	0	0	0	0	8
	sociolinguistique	8	0	0	0	0	0	8
	étendue	0	8	0	0	0	0	8
	grammaire	8	0	0	0	0	0	8
	cohérence	8	0	0	0	0	0	8

Résultats des votes des experts au critère adéquation

Votes	Critère	Oui	Non	Totaux
individuels	adéquation	0	8	8
après discussion	adéquation	0	8	8

Cette production est de niveau A1.1-.

3.4. Copie 4

Avant et après le travail de groupe, les experts n'ont pas évalué le critère *cohérence*. Tous les experts estiment que la production est de niveau global A1.1-. Le niveau varie selon les critères. La plupart des experts accordent le niveau A1.1- au critère *sociolinguistique*, le niveau A1.1+ au critère *étendue* et le niveau A1 au critère *grammaire*.

Avant et après la discussion, tous les experts pensent que le candidat n'a pas respecté la consigne de l'exercice (critère *adéquation*).

Résultats des votes des experts

Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
Individuels	note	8	0	0	0	0	0	8
	sociolinguistique	8	0	0	0	0	0	8
	étendue	0	0	7	1	0	0	8
	grammaire	0	1	0	7	0	0	8
	cohérence	0	0	0	0	0	0	0
Après discussion	note	8	0	0	0	0	0	8
	sociolinguistique	8	0	0	0	0	0	8
	étendue	0	0	8	0	0	0	8
	grammaire	0	0	0	8	0	0	8
	cohérence	0	0	0	0	0	0	0

Résultats des votes des experts au critère adéquation

Votes	Critère	Oui	Non	Totaux
individuels	adéquation	0	8	8
après discussion	adéquation	0	8	8

Cette production est de niveau A1.1-.

3.5. Copie 5

Lors du vote individuel, plus de la moitié des experts pensent que la production est de niveau A1.1 globalement ainsi qu'aux 4 critères.

Après le travail de groupe, tous les experts estiment que la production est de niveau A1.1 globalement et à tous les critères.

Avant et après la discussion, tous les experts pensent que le candidat a respecté la consigne de l'exercice (critère *adéquation*).

Résultats des votes des experts

Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
Individuels	note	0	8	0	0	0	0	8
	sociolinguistique	0	4	0	3	0	0	7
	étendue	0	6	0	2	0	0	8
	grammaire	2	6	0	0	0	0	8
	cohérence	2	5	0	1	0	0	8
Après discussion	note	0	8	0	0	0	0	8
	sociolinguistique	0	8	0	0	0	0	8
	étendue	0	8	0	0	0	0	8
	grammaire	0	8	0	0	0	0	8
	cohérence	0	8	0	0	0	0	8

Résultats des votes des experts au critère adéquation

Votes	Critère	Oui	Non	Totaux
individuels	adéquation	8	0	8
après discussion	adéquation	8	0	8

Cette production est de niveau A1.1.

3.6. Copie 6

Lors du vote individuel, 5 experts pensent que la production est de niveau A1.1 globalement. Les trois quarts estiment qu'elle est de niveau A1.1- au critère *grammaire* et de niveau A1 pour la *cohérence*. Tous les experts attribuent le niveau A1 au critère *sociolinguistique*. Leur avis est partagé entre les niveaux A1.1 et A1 pour le critère *étendue*.

Après le travail de groupe, tous les experts estiment que la production est de niveau A1.1 globalement, ainsi qu'aux critères *étendue* et *grammaire*. Les 8 experts attribuent le niveau A1 aux critères *sociolinguistique* et *cohérence*.

Avant et après la discussion, tous les experts pensent que le candidat a respecté la consigne de l'exercice (critère *adéquation*).

Résultats des votes des experts

Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
Individuels	note	2	5	0	1	0	0	8
	sociolinguistique	0	0	0	8	0	0	8
	étendue	0	4	0	4	0	0	8
	grammaire	6	2	0	0	0	0	8
	cohérence	2	0	0	6	0	0	8
Après discussion	note	0	8	0	0	0	0	8
	sociolinguistique	0	0	0	8	0	0	8
	étendue	0	8	0	0	0	0	8
	grammaire	0	8	0	0	0	0	8
	cohérence	0	0	0	8	0	0	8

Résultats des votes des experts au critère adéquation

Votes	Critère	Oui	Non	Totaux
individuels	adéquation	8	0	8
après discussion	adéquation	8	0	8

Cette production est de niveau A1.1.

3.7. Copie 7

Lors du vote individuel, la majorité des experts attribuent le niveau A1.1 globalement à la production ainsi qu'aux critères *grammaire* et *cohérence*. Tous les experts donnent le niveau A1 au critère *cohérence*. Leur avis est partagé entre les niveaux A1.1- et A1.1 pour le critère *sociolinguistique*.

Après le travail de groupe, tous les experts attribuent à la production le niveau A1.1 globalement, ainsi qu'aux critères *cohérence*, *sociolinguistique*, et *grammaire*. Les 8 experts pensent qu'elle est de niveau A1 au critère *étendue*.

Avant et après la discussion, tous les experts pensent que le candidat a respecté la consigne de l'exercice (critère *adéquation*).

Résultats des votes des experts

Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
Individuels	note	0	8	0	0	0	0	8
	sociolinguistique	4	4	0	0	0	0	8
	étendue	0	2	0	6	0	0	8
	grammaire	0	7	0	0	0	0	7
	cohérence	0	5	0	1	0	0	6
Après discussion	note	0	8	0	0	0	0	8
	sociolinguistique	0	8	0	0	0	0	8
	étendue	0	0	0	8	0	0	8
	grammaire	0	8	0	0	0	0	8
	cohérence	0	8	0	0	0	0	8

Résultats des votes des experts au critère adéquation

Votes	Critère	Oui	Non	Totaux
individuels	adéquation	8	0	8
après discussion	adéquation	8	0	8

Cette production est de niveau A1.1.

3.8. Copie 8

Que ce soit avant ou après la discussion, tous les experts pensent que cette production est de niveau A1.

Lors du vote individuel, au moins trois quart des experts attribuent le niveau A1 aux critères *cohérence*, *sociolinguistique* et *grammaire*. Leur avis est partagé entre les niveaux A1.1 et A1 pour le critère *étendue*.

Après le travail de groupe, tous les experts estiment que la production est de niveau A1 globalement ainsi qu'aux 4 critères.

Avant et après la discussion, tous les experts pensent que le candidat a respecté la consigne de l'exercice (critère *adéquation*).

Résultats des votes des experts

Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
Individuels	note	0	0	0	8	0	0	8
	sociolinguistique	0	0	0	8	0	0	8
	étendue	0	4	0	4	0	0	8
	grammaire	0	1	0	6	1	0	8
	cohérence	0	2	0	6	0	0	8
Après discussion	note	0	0	0	8	0	0	8
	sociolinguistique	0	0	0	8	0	0	8
	étendue	0	0	0	8	0	0	8
	grammaire	0	0	0	8	0	0	8
	cohérence	0	0	0	8	0	0	8

Résultats des votes des experts au critère adéquation

Votes	Critère	Oui	Non	Totaux
individuels	adéquation	8	0	8
après discussion	adéquation	8	0	8

Cette production est de niveau A1.

3.9. Copie 9

Avant et après le travail de groupe, les 8 experts estiment que cette production est de niveau A1 globalement ainsi qu'aux 4 critères.

Avant et après la discussion, tous les experts pensent que le candidat a respecté la consigne de l'exercice (critère *adéquation*).

Résultats des votes des experts

Votes	Critères	A1.1-	A1.1	A1.1+	A1	A1+	A2	Totaux
Individuels	note	0	0	0	8	0	0	8
	sociolinguistique	0	0	0	8	0	0	8
	étendue	0	0	0	8	0	0	8
	grammaire	0	0	0	8	0	0	8
	cohérence	0	0	0	8	0	0	8
Après discussion	note	0	0	0	8	0	0	8
	sociolinguistique	0	0	0	8	0	0	8
	étendue	0	0	0	8	0	0	8
	grammaire	0	0	0	8	0	0	8
	cohérence	0	0	0	8	0	0	8

Résultats des votes des experts au critère adéquation

Votes	Critère	Oui	Non	Totaux
individuels	adéquation	8	0	8
après discussion	adéquation	8	0	8

Cette production est de niveau A1.

Conclusion

Les votes des experts sont plus homogènes après la discussion en groupe. Nous obtenons, en nous basant sur les votes après discussion, les niveaux suivants pour chaque copie de production écrite :

Niveaux des productions

Copie	Laisser un message
1	A1.1-
2	A1.1-
3	A1.1-
4	A1.1-
5	A1.1
6	A1.1
7	A1.1
8	A1
9	A1



Annexes





Kappa pour plusieurs juges et plusieurs catégories (Fleiss)

Les calculs du coefficient kappa ont été effectués à l'aide du logiciel R, package *irr* (fonction *kappam.fleiss*).

Soit m juges qui doivent classer n observations dans k catégories ($m > 2$ et $k > 2$).

Le coefficient kappa peut être défini de la façon suivante :

$$K = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}$$

Pour la $j^{\text{ème}}$ catégorie :

m_{ij} : nombre de juges qui ont mis la $i^{\text{ème}}$ observation dans la $j^{\text{ème}}$ catégorie.

$$p_j = \frac{1}{nm} \sum_{i=1}^n m_{ij}$$

$$q_j = 1 - p_j$$

$$P_i = \frac{1}{m(m-1)} \sum_{j=1}^k m_{ij}(m_{ij} - 1)$$

Pourcentage de paire de juges en accord :

$$\bar{P} = \frac{1}{n} \sum_{i=1}^n P_i$$

Pourcentage de pair de juges jugeant de façon aléatoire

$$\bar{P}_e = \sum_{j=1}^k p_j^2$$

Sous l'hypothèse des jugements aléatoire, nous retrouvons que $k \sim N(0, \bar{s}_0)$:

$$\bar{s}_0 = \frac{\sqrt{2}}{\sum_{j=1}^k p_j q_j \sqrt{nm(m-1)}} \sqrt{\left(\sum_{j=1}^k p_j q_j \right) - \sum_{j=1}^k p_j q_j (q_j - p_j)}$$